

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:15:45 ; Search time 23.19 Seconds
(without alignments)
887.985 Million cell updates/sec

Title: US-09-866-034-2
Perfect score: 1399
Sequence: 1 MHKAGLLGICARAMNSVRMA.....DASYINGEYVVGCTPSRL 278

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq..1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1399	100.0	278	21	PRO1800, a Hep27 h
2	1395	99.7	278	22	AAV96729
3	1327	94.9	278	21	AAV68735
4	649.5	46.4	183	21	AAV58822
5	588.5	42.1	254	21	AAV41219
6	588.5	42.1	276	21	AAV41218
7	428	30.6	103	22	AAV74580
8	404	28.9	248	21	AAV54422
9	370	26.4	79	21	AAV54422
10	353	25.2	262	21	AAV54422
11	349	24.9	292	22	AAV54422

12	349	24.9	295	22	AAV79400	Corynebacterium g1
13	349	24.9	295	22	AAV79401	Corynebacterium g1
14	348	24.9	246	21	AAV15707	Staphylococcus aur
15	348	24.9	246	22	AAE02195	S. aureus NADPH-de
16	345	24.7	253	21	AAV31819	Arabidopsis thalia
17	341	24.4	243	22	AAV01032	Arabidopsis thalia
18	340	24.3	243	19	AAV80670	CPE 35 protein seq
19	340	24.3	243	19	AAV80670	S. pneumoniae fat
20	339	24.2	254	22	AAV49773	Streptococcus pneu
21	339	24.2	254	22	AAV49773	Protein with acety
22	336	24.0	306	22	AAV81644	S. epidermidis ope
23	333.5	23.8	251	20	AAV36959	Chlamydia trachoma
24	333.5	23.8	251	20	AAV34891	S. epidermidis ope
25	331.5	23.7	261	21	AAV96271	Chlamydia pneumoni
26	331.5	23.7	261	21	AAV96271	S. epidermidis ope
27	331.5	23.7	261	21	AAV96271	B. subtilis glucos
28	330	23.6	263	21	AAV54424	Amino acid sequenc
29	330	23.6	271	21	AAV43271	Arabidopsis thalia
30	328.5	23.5	315	17	AAV43270	Arabidopsis thalia
31	328.5	23.5	315	17	AAV43270	Arabidopsis thalia
32	327.5	23.4	340	21	AAV10740	Arabidopsis thalia
33	327.5	23.4	340	21	AAV10741	Rape leaf beta-ket
34	326.5	23.3	261	13	AAV27757	Rape seed beta-ket
35	326.5	23.3	261	13	AAV27757	B. megaterium gluc
36	325.5	23.3	279	21	AAV49315	H. ghiliani/B. me
37	325.5	23.3	261	9	AAV80590	Glucose dehydrogen
38	324	23.2	262	13	AAV27756	Arabidopsis thalia
39	324	23.2	262	13	AAV27756	Sequence of glucos
40	324	23.2	262	22	AAV86337	NAD affinity glucos
41	323.5	23.1	271	21	AAV08828	ylitol dehydrogen
42	323.5	23.1	254	21	AAV32505	G. suboxydans DSM
43	323.5	23.1	264	21	AAV49317	Arabidopsis thalia
44	323.5	23.1	269	21	AAV49316	Arabidopsis thalia
45	323.5	23.1	318	21	AAV29313	Arabidopsis thalia
			322	21	AAV29312	Arabidopsis thalia

ALIGNMENTS

RESULT	1	AAV96729	standard; Protein: 278 AA.
ID	AAV96729	standard; Protein: 278 AA.	
AC	AAV96729;		
XX	26-SEP-2000	(first entry)	
DE	PRO1800, a Hep27 homologue.		
XX	PRO1800; Hep27; homologue; short-chain alcohol dehydrogenase; SCAD;		
KW	secreted protein; transmembrane protein; recombinant production;		
KW	gene therapy.		
XX			
OS	Homo sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	Peptide	1..15	
FT	Modified-site	43..48	/label= Signal_peptide
FT	Modified-site	/note= "N-myristoylation site"	
FT	Modified-site	80..85	/note= "N-myristoylation site"
FT	Modified-site	/note= "N-myristoylation site"	
FT	Modified-site	183..186	/note= "N-glycosylation site"
FT	Modified-site	191..196	/note= "N-glycosylation site"
FT	Modified-site	213..218	/note= "N-myristoylation site"
FT	Modified-site	272..277	/note= "N-myristoylation site"
FT	Modified-site	276..278	/note= "N-myristoylation site"
FT	Peptide	276..278	/label= Targeting_signal
FT		/note= "Microbody C-terminal targeting signal"	

XX WO200036102-A2.
 XX 22-JUN-2000.
 XX 01-DEC-1999; 99WO-US28634.
 XX 16-DEC-1998; 98US-0112851.
 XX 16-DEC-1998; 98US-0113145.
 XX 22-DEC-1998; 98US-0113511.
 XX 12-JAN-1999; 99US-0115558.
 XX 12-JAN-1999; 99US-0115565.
 XX 12-JAN-1999; 99US-0115733.
 XX 09-FEB-1999; 99US-0119341.
 XX 10-FEB-1999; 99US-0119537.
 XX 12-FEB-1999; 99US-0119965.
 XX 02-JUN-1999; 99WO-US12252.
 XX (GETH) GENENTECH INC:
 XX Botstein D, Desnovers L, Ferrara N, Fong S, Gao W, Goddard A;
 PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
 PI Wood WI;
 XX WPI: 2000-431586/37.
 DR N-PSDB; AANA51259.

PT Isolated nucleic acid molecule encodes a PRO polypeptide which is a
 PT transmembrane polypeptide

PS Claim 12; Fig 2; 154pp; English.

XX This is PRO1800, a putative human Hep27 protein homologue, which has
 CC recently been shown to have homology to known short-chain alcohol
 CC dehydrogenase (SCAD) family of proteins. The invention concerns novel
 CC secreted and transmembrane proteins, designated PRO polypeptides. The
 CC cDNA and gene sequences are useful in the recombinant production of PRO
 CC polypeptides, as a hybridization probe to screen libraries to isolate
 CC cDNAs with sequence identity to PRO polypeptides or to map the gene
 CC encoding the PRO polypeptides and analyzing genetic disorders. The
 CC cDNA/gene can also be used to produce transgenic animals useful for the
 CC development and screening of therapeutically useful reagents. They can
 CC also be used in gene therapy, e.g. to replace a defective gene.

SQ Sequence 278 AA;

Query Match 100.0%; Score 1399; DB 21; Length 278;
 Best Local Similarity 100.0%; Pred. No. 4,9e-132;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHKAGLLGLCARAWNSYRMASSGMRTRDPLANKVALVTASDGIQFARLAQDGAHV 60
 Db 1 mhkagllglcarawnsyrmasgmrtrdplankvalvtascdgiqfatarlaqdgahv 60
 QY 61 VSSKKQONDOAVATLOGEGLSVTVCHWGKADRERLVATAVAKLGIDILVSNAVN 120
 Db 61 vsskkqonvdavatlqeglsvtgchvgykaedrervlatavaklkgidilvsnaavn 120
 QY 121 PFFCSINDVTEWMDKLDINVKAPALMTKAVVPEMKRGGSVYVSSIAESPSGFS 180
 Db 121 pffcsindvteewmdkldinvkapaalmkavvpemkrggsvyvsiaaefspgfs 180
 QY 181 PYNVSKTALLGLTKTLAIELAPRINIRVNCIAPGLIKTSFSRMLMDKEESKKTIR 240
 Db 181 pynvsktallgltktlaielaprinirvnciapglktsfsmldmdekeskktlir 240
 QY 241 RLSEPECCATVSLFCSEDAVYITGETVNVGGTSPRL 278
 Db 241 rlsepedcagivslfcsedasyitgetvnyvggtsprrl 278

RESULT 2

AAB93414
 ID AAB93414 standard; Protein; 278 AA.
 XX AAB93414;
 AC 26-JUN-2001 (first entry).
 DT Human protein sequence SEQ. ID NO:12620.
 DE Human protein sequence SEQ. ID NO:12620.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.

PN EPI074617-A2.
 PD 07-FEB-2001.
 PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 XX WPI: 2001-316749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 12620; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide comprising a sequence
 CC complementary to the complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SQ Sequence 278 AA;

Query Match 99.7%; Score 1395; DB 22; Length 278;
 Best Local Similarity 99.6%; Pred. No. 1.2e-131;
 Matches 277; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHKAGLLGLCARAWNSYRMASSGMRTRDPLANKVALVTASDGIQFARLAQDGAHV 60
 Db 1 mhkagllglcarawnsyrmasgmrtrdplankvalvtascdgiqfatarlaqdgahv 60

FT	Region	/note= "SCAD family signature sequence"
FT		182..201
FT		/note= "glucose/ribitol dehydrogenase motif"
FT	Domain	187..186
FT		/note= "canonical catalytic site of SCADs"
FT	Region	203..220
FT		/note= "glucose/ribitol dehydrogenase motif"
FT	Region	204..213
FT		/note= "SCAD motif"
FT	Modified-site	232
FT		/note= "potential casein kinase II phosphorylation site and potential protein kinase C phosphorylation site"
FT	Modified-site	236
FT		/note= "potential protein kinase C phosphorylation site"
FT	Region	238..258
FT		/note= "glucose/ribitol dehydrogenase motif"
PX		
PX	WO200004135-A2.	
PD		
PE	27-JAN-2000.	
PF		
PR	16-JUL-1999;	99WO-US16164.
PR	16-JUL-1998;	98US-O116750.
PR	16-JUL-1998;	98US-O160074.
PA	(INCY-) INCYTE PHARM INC.	
PL	Bandman O, Tang YT, Corley NC, Azimzai Y, Baughn MR;	
DR	WPI: 2000-171266/15.	
XX	N-PSDB; AAZ460930.	
PT	New short chain alcohol dehydrogenase polypeptides useful for diagnosis, treatment and prevention of cell proliferative disorders such as atherosclerosis, cirrhosis and cancers of various tissues -	
XX	Claim 1; Fig 1A-D: 78pp; English.	
CC	The present sequence represents a human short chain alcohol dehydrogenase (SCAD)-related molecule designated SCR-M-1. The specification also describes SCR-M-2. SCR-M proteins are metabolic, cell proliferation and inflammation regulators. The SCR-M polynucleotides and polypeptides are used for treating or preventing a cell proliferative or immune disorder in humans. Cell proliferative disorders include arteriosclerosis, atherosclerosis, psoriasis, cirrhosis, and hepatitis. Immune disorders include AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, and amyloidosis. The vectors, agonists, antagonists, antibodies and complementary sequences are also used for treating the above conditions. The polynucleotides and polypeptides are also used for treating cancers of various tissues such as adrenal gland, bladder, bone, bone marrow, and brain.	
XX	Sequence	278 AA:
SQ		
Query Match	94.9%;	Score 1327; DB 21; Length 278;
Best Local Similarity	96.4%;	Fred. No. 8.2e-125;
Matches 268;	Conservative 0;	Mismatches 10; Indels 0; Gaps 0
OY	1 MHKAGILGLCARAMNSVRRASSGMRPDLAKKVALVTASDGGFARLADODGAHV	60
Db	1 Mhmarrlllgcawksrsvrmassmrrlrlplntkvalvaascdgigfataarllagdrhhv	60
OY	61 VSSRKQGNVDQAVATLOGEGLSVTGTCHVGCAEDREFLVATAVKLNGIDITVSNAVN	120
Db	61 vssrkqgnvdqavatlgeglsvtgcvhvgaedrerelvatavklngiditlvsnavn	120
OY	121 PFGSGINDTEEFNPKTIDINKAPALMTKANVPENMERGGGSVIYTSIAAFSPSGFS	180
Db	121 pffgsgindteefnwktidlnkapalmtkavpenmekrggsvviytsiaafspsgfs	180

QY 181 PYNVSKTALLGLTTLAIELAPRIRVNCIAPGLIKTSFSRMLMDKEKESMKETLRIR 240
 DB 181 pynvsktallgltnltaielaprirvnciapgliktsfsmldmkekesmktlrlir 240
 QY 241 RLGEPEDCAGIVSFLCEDASYIGETVWVGCGPPSRU 278
 DB 241 rlgepedcagivsficesdasylgetvwwvggcpssrl 278

RESULT 4
 ID AAB58822 standard; Protein; 183 AA.
 AC AAB58822;
 DT 27-MAR-2001 (first entry)
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 530.

Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 neurotropic; neuroprotective; antiviral; antiinflammatory; hepatotropic;
 antiinflammatory; antiinflammatory; antitumor; antitumor; antitumor;
 antibacterial; antifungal; antiparasitic; candidant; immune disorder;
 Addison's disease; allergy; autoimmune haemolytic anaemia;
 autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.
 OS
 XX
 PN MO200055173-A1.
 PL 21-SEP-2000.
 PF 08-MAR-2000; 2000MO-US05881.
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 DR MPI: 2000-611515/58.
 DR N-PSDB; AAF21175.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11; Page 963-964; 1299pp; English.
 XS
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiinflammatory; hepatotropic;
 CC antibacterial; antiinflammatory; antitumor; antitumor; antitumor;
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and agonists may also be used in the diagnosis of cancer,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 XX Sequence 183 AA;
 SQ

Query Match 46.4%; Score 649.5; DB 21; Length 183;
 Best Local Similarity 55.2%; Pred. No. 4e-57;
 Matches 148; Conservative 2; Mismatches 21; Indels 97; Gaps 3;

QY 16 SVRMASSGMRDRDPLANKVALVTASTDIGTAIRRLAODGAHVVSRRKQNDQAV-- 73
 DB 8 svrmasstrldpblnkvalvtastdigf-----aspvwprrtprgrgdeaeecg 61
 QY 74 ---ATLOGELSVGTVCYCHVKAEDRELVATAVKLGCGIDILVSNAAVNFESINDYT 130
 DB 62 psggtlqgeglsvtgctcxxykaedrelvatavklingidilvsnaavpflgslndvt 121
 QY 131 EEVWDKTLIDINVKAPALMTKAVPEMEKRGGSVIVSSIAFSPSPGFSFYNVSKTALL 190
 DB 122 eevwdk----- 127
 QY 191 GLTTLAIELAPRIRVNCIAPGLIKTSFSRMLMDKEKESMKETLRIRLGEPEDCAG 250
 DB 128 -----lmdkekesmktlrlrlgepedcag 155
 QY 251 IVSFLCEDASYIGETVWVGCGPPSRU 278
 DB 156 ivsficesdasylgetvwwvggcpssrl 183

RESULT 5
 ID AAG41219 standard; Protein; 254 AA.
 AC AAG41219;
 DT 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 51256.
 XX
 XX Protein identification: signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 OS
 XX
 PN EP1033405-A2.
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135121.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151338.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

Query Match	Score 588.5;	DB 21;	Length 276;
Best Local Similarity	50.6%;	Pred. No. 9,5e-51;	
Matches 126;	Conservative 35;	Mismatches 85;	Indels 3; Gaps 2
Pr 21-OCT-1999;	99US-0160741		
Pr 21-OCT-1999;	99US-0160767		
Pr 21-OCT-1999;	99US-0160768		
Pr 21-OCT-1999;	99US-0160770		
Pr 21-OCT-1999;	99US-0160814		
Pr 21-OCT-1999;	99US-0160815		
Pr 22-OCT-1999;	99US-0160980		
Pr 22-OCT-1999;	99US-0160981		
Pr 25-OCT-1999;	99US-0160989		
Pr 25-OCT-1999;	99US-0161404		
Pr 25-OCT-1999;	99US-0161405		
Pr 26-OCT-1999;	99US-0161406		
Pr 26-OCT-1999;	99US-0161359		
Pr 26-OCT-1999;	99US-0161360		
Pr 26-OCT-1999;	99US-0161361		
Pr 28-OCT-1999;	99US-0161920		
Pr 28-OCT-1999;	99US-0161992		
Pr 28-OCT-1999;	99US-0161993		
Pr 29-OCT-1999;	99US-0162142		
Query Match	42.1%;	Score 588.5;	DB 21;
Best Local Similarity	50.6%;	Pred. No. 9,5e-51;	
Matches 126;	Conservative 35;	Mismatches 85;	Indels 3; Gaps 2
Pr 30 LANKALVLTASDGCISFAIRRLADGAHVVSRRQNVDAVATLQEGSLVTGYCH 89			
Pr 31 Lsgkvaivaivastqigjfiglterfiglegasavvsrrkqanvdeavaklksgjdaiyivch 90			
Pr 90 VCKAEDRELRVLTAVKTLHGGCIDILVSNAAVNPFFGSIIMVTEWWDKTLIDIVKAPALMT 149			
Pr 91 vnaagtrlnrlivekvtqkygkidlvcnaaampstldpllskkaavidklwehivkssllll 150			
Pr 150 KAVVEMERKGGGVYVTSIAAFSPSPGFSPYNSKTTALLGLTTLTALTELAPRNRVNC 209			
Pr 151 qdmaphlek-gssvlfitsisagfpqgamamyvtktalligltaklaaemap-dtrvna 207			
Pr 210 LAPGLIKTSFSFMLMNDKKEESMKETLIRRLGEGEDCAGIVSFLCSEDASYINGETVY 269			
Pr 208 vavgfvtptfiastflgssevregieektlrnrlgtgtmdnaaaafaaadssysltgetlv 267			
Pr 270 VGGTSPSRL 278			
Pr 268 vagampstl 276			
RESULT 7			
AAAG74580			
ID AAAG74580. standard; Protein: 103 AA.			
XX AC	AAAG74580;		
XX DT	03-SEP-2001 (first entry)		
XX DE	Human colon cancer antigen protein SEQ ID NO:5344.		
XX KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
XX KW	colorectal carcinoma.		
XX OS	Homo sapiens.		
XX PD	MO20012920-A2.		
XX PE	05-APR-2001.		
XX PR	28-SEP-2000; 2000WO-US26524.		
XX PR	29-SEP-1999; 99US-0157137.		
XX PR	03-NOV-1999; 99US-0163280.		
XX PA	(HUMA-) HUMAN GENOME SCI INC.		
XX TI	Ruben SM, Barash SC, Birse CE, Rosen CA.		

DR WPI: 2001-235357/24.
 DR N-PSDB; AAH33985.
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11: Page 6991-6992; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 103 AA;
 Query Match 30.6%; Score 428; DB 22; Length 103;
 Best Local Similarity 94.6%; Pred. No. 2.8e-35;
 Matches 87; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MHKAGLLGICARWNSVRNASSGRTTRDPLANKVALVTASTDTCIGFARLADGAHV 60
 DB 2 mhkaagllgicaratamsvrnasssgmrttrdplankvalvtastdgcgafarladgahv 61
 QY 61 VSSRKOONVQAVATLOGEGLSVTGTVCHYK 92
 DB 62 vssrkoonvqavatlqgeglsvtgvcamwgr 93
 RESULT 8
 AAY54422
 ID AAY54422 standard; Protein: 248 AA.
 AC AAY54422;
 DT 06-APR-2000 (first entry)
 XX
 DE Amino acid sequence of a beta-ketoacyl-ACP reductase protein.
 XX
 KW Beta-ketoacyl-ACP reductase: Type II fatty acid synthetase;
 KW stereospecificity; 4-chloroacetoacetic acid ester;
 KW (S)-4-halo-3-hydroxybutyric acid ester; 4-halo-acetoacetic acid ester;
 KW beta-ketoacyl-acyl carrier protein reductase; acetoacetyl-CoA reductase;
 KW polybeta-hydroxy fatty acid biosynthesis; optically active;
 KW 4-haol-3-hydroxybutyric acid ester.
 XX
 OS Bacillus subtilis.
 XX
 PN EP955375-A2.
 XX
 PD 10-NOV-1999.
 XX
 XX 10-MAY-1999; 99EP-0109403.
 XX
 XX 08-MAY-1998; 98JP-0126507.
 PR 21-OCT-1998; 98JP-0300178.
 PR 05-APR-1999; 99JP-0098205.
 XX

PA (DAIL) DAICEL CHEM IND LTD.
 XX
 PI Yamamoto H;
 XX
 DR WPI: 2000-118183/11.
 DR N-PSDB; AA245749.
 XX
 PT Producing optically pure (S)-4-halo-3-hydroxybutyric acid ester -
 XX
 PS Claim 6; Page 19-20; 34pp; English.
 XX
 CC The present sequence represents a beta-ketoacyl-ACP reductase protein
 CC of Bacillus subtilis. The beta-ketoacyl-ACP reductase enzyme constitutes
 CC a type II fatty acid synthetase. The enzyme has an extremely high
 CC reducing activity and stereospecificity towards 4-chloroacetoacetic
 CC acid ester. The specification describes a method for producing a
 CC (S)-4-halo-3-hydroxybutyric acid ester. The method comprises
 CC asymmetrically reducing 4-halo-acetoacetic acid ester or its
 CC derivative with beta-ketoacyl-acyl carrier protein reductase
 CC constituting Type II fatty acid synthase, or acetoacetyl-CoA
 CC reductase constituting the polybeta-hydroxy fatty acid biosynthesis
 CC system. The novel method is used to produce optically active
 CC 4-haol-3-hydroxybutyric acid ester, with a high purity.
 XX
 SQ Sequence 248 AA;
 Query Match 28.9%; Score 404; DB 21; Length 248;
 Best Local Similarity 37.7%; Pred. No. 2.6e-32;
 Matches 93; Conservative 50; Mismatches 100; Indels 4; Gaps 3;
 QY 28 DPLANKVALVTASTDTCIGFARLADGAHVVS-SRKOONVQAVATLOGEGLSVTGT 86
 DB 2 dmlndkcalvtgsgvgrlslalalaksaganvvnvysgneakanevdelksmgrikav 61
 QY 87 VCHVGAEDRERLVAFVAKLHGIDILVSNAAVNPFGSITMDVTEWMDTLDINKAPA 146
 DB 62 kadvspedvgmnikelvsfstcidilvnmagltr-dnlmrxkdedwdvlnlkyvf 120
 QY 147 LMRKAVPEMEKRGSGSVIVSSIAAFSPSPGFSYVNSKTALLGLTKTLAELAPRNR 206
 DB 121 nckkavtrgmkkqgrsgrlsvssilvysgngpganyaaakagylgltksakelaarnt 180
 QY 207 VNCLARGLITFSRMLMNDKEEKESKFTLRRLGEPEDCAGIVSFLCEDASYITGE 266
 DB 181 vnlaipgflstmdtkl--akdvqdemlkqiplarfgepdsvsyvtflasegarymtg 238
 QY 267 TVVWGGG 273
 DB 239 tlnhdgg 245.
 RESULT 9
 AAG02114
 ID AAG02114 standard; Protein: 79 AA.
 AC AAG02114;
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 6195.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 XX 06-SEP-2000.
 PD 21-FEB-2000; 2000EP-0200610.
 XX
 XX

```

PR 26-FEB-1999; 9905-0122487.
XX
XX (GEST ) GENSET.
PI Dumas Mline Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
DR N-PSDB; AAC02120.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 6195; 71pp + CD-ROM; English.
PS
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs with intact 5'
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
CC
XX Sequence 79 AA:
SO

```

Query Match 26.4%; Score 370; DB 21; Length 79;
 Best Local Similarity 96.2%; Pred. No. 1.3e-29;
 Matches 76; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MHKAGLGLICARANMSVMASGMRDPLANKVALVTASTDGIFFAIRLAODCAHYV 60
Db 1 mxaxxllglcarawsvmassgmrrtriplankvalvtastdgiffalarlaodgahyv 60
QY 61 VSSRKQNWDDAVATLQGE 79
Db 61 vstrkqnvddavatlqge 79

```

RESULT 10
 AAG31818
 ID AAG31818 standard; Protein: 262 AA.
 XX
 AC AAG31818;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 38275.
 XX
 KM Protein identification: signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 9905-0121825.
 PR 05-MAR-1999; 9905-0123180.
 PR 09-MAR-1999; 9905-0123548.
 PR 23-MAR-1999; 9905-0125788.
 PR 25-MAR-1999; 9905-0126264.
 PR 29-MAR-1999; 9905-0126785.

```

PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 14-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139753.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142380.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.

```

Query Match	25.2%;	Score 353;	DB 21;	Length 262;
Best Local Similarity	35.0%;	Pred. No. 3.7e-27;		
Matches 89;	Conservative 47;	Mismatches 110;	Indels 8;	Gaps 3

Sequence 292 AA;

malcines 80; conservative 45; Mismatches 114; Indels 4; Gaps 2;

Db 105 qdlrdnevrslvaetyna]ag]dli]nnaacromu]to]it]de]dd]t].....

208 NCLAPGLIKTSF SRMLWMDKEKEESMKETLRIRLGEPEDCAGIVSFLCSEDA
SYITGET 267

Corynebacterium glutamicum.

PN WO200100844-A2

PD 04-JAN-2001

PF 23-JUN-2000: 2000WO-TB00943

25-TTN-1999. 90115-0141037

PR 08-JUL-1999; 99DE-1031412.
 BR 08-JUL-1999; 99DE-1031413

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031431.

08-JUL-1999: 99DE-1031434

08 JUL-1999; 99DE-1031510;
08-JUL-1999; 99DE-1031510;
PR

PK 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032973

PR 27-AUG-1999: 99DF-10A0765

PR	03-SEP-1999	98DE-1013035
JK	01 AUG 1999	99US-0131572

PR 03-SEP-1999; 99DE-1042079.
03-SEP-1999 0057 1042079

03-SEP-1999; 99DE-1042087.

PR 03-SEP-1999; 99DE-1042095.

PR 03-SEP-1999; 99DE-1042125

PA (BADI) BASE AG

Downloaded from www.sagepub.com at UNIV OF CALIF SDIEN on May 12, 2015

XX XX

DR N-PSDB; AAF71517.

PT New isolated *Corynebacterium*

modulation of production of

carbohydrates or enzymes -

Claim 20; Page 585-586; 1246p

AAF71360 to AAF71750 encode

CC AAB79243 to AAB 79633 which are

(II) for expression in host cell

or nonproteinogenic amino acids.

acid nucleoside, a nucleotide, a

cofactor, a polyketide, or an

CC Corynebacterium diphtheriae

Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells

Query Match	24.9%	Score	349	DB	22	Length	295
Best Local Similarity	34.5%	Pred. No.	1.1e-26				
Matches	86	Conservative	45	Mismatches	114	Indels	4
						Gaps	2

QY		30	LANKVAIVTASTDDEIGFAIRRLAODDAHVVS--SRKOOONVDDAVATLOGEGLSVTGYV	87
Dd		48	lkjrkalltggdsiglgaavalaayaregadvaiaimlpeeagadadtvljaieetqkafsfp	107
QY		88	CHVCKAEDRELVATAVKLHGGLIDILVSNAAVNEFFGSINDYTEEVDKTLIDINVKAPAL	147
Dd		108	gdIrdpeyrcaislygetvnaIlgldlllvnnasrygwapglteitdentdgclgynylysfr	167
QY		148	MTRKAVPEEMKEKRGSGVVVISSTAESPSPGFSRYNVNSKTALLGTYTLIELAPRIRIV	207
Dd		168	vtkaaiphl--kpssstlfssiyqaypssetlIdyaamtkaalmlslqlasslIggdriv	225
QY		208	NCLAPGLIKTSFSLMLWMDREKESMEKETLRIRLGEPEDCAGIVSFLCSEDASYRTGET	267
Dd		226	nsvaypgfwfcpIqpsbhgppqeiklegfgqhpnIgragmpvelajagayvllaadesyyvvet	285
QY		268	VWVGCGTPS 276	
Dd		286	Igvrgtfcpt 294	

RESULT	13
AAB79401	
ID	AAB79401 standard; protein; 295 AA.
XY	

DT 30-APR-2001 (first entry)
XY

Corynebacterium glutamicum SMP protein sequence SEQ ID NO:318.

KM Corynebacterium glutamicum; carbon metabolism and energy production;
KM SMP protein; sugar metabolism and oxidative phosphorylation protein;
KM fine chemical production; organic acid; proteino-genic amino acid;
KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KM carbohydrate; aromatic compound; Vitamin; cofactor; polyketide; enzyme
diagnosis; Corynebacterium diptheriae; evolutionary study.

05 *Corynebacterium glutamicum*.
yy

PN WO200100844-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-IB00943.

PR	25-JUN-1999;	99US-1041031
PR	08-JUL-1999;	99DE-1031412
PR	08-JUL-1999;	99DE-1031413
PR	08-JUL-1999;	99DE-1031419
PR	08-JUL-1999;	99DE-1031420
PR	08-JUL-1999;	99DE-1031424
PR	08-JUL-1999;	99DE-1031428
PR	08-JUL-1999;	99DE-1031431
PR	08-JUL-1999;	99DE-1031431
PR	08-JUL-1999;	99DE-1031433
PR	08-JUL-1999;	99DE-1031510
PR	08-JUL-1999;	99DE-1031562

PR	08-JUL-1999:	99DE-1031634.
PR	09-JUL-1999:	99DE-1032180.
PR	09-JUL-1999:	99DE-1032227.
PR	09-JUL-1999:	99DE-1032230.
PR	09-JUL-1999:	99DS-0143208.
PR	14-JUL-1999:	99DE-1032974.
PR	14-JUL-1999:	99DE-1032973.
PR	14-JUL-1999:	99DE-1033005.
PR	27-AUG-1999:	99DE-1040765.
PR	31-AUG-1999:	99DS-0151572.
PR	03-SEP-1999:	99DE-1042079.
PR	03-SEP-1999:	99DE-1042086.
PR	03-SEP-1999:	99DE-1042087.
PR	03-SEP-1999:	99DE-1042088.
PR	03-SEP-1999:	99DE-1042095.
PR	03-SEP-1999:	99DE-1042123.
PR	03-SEP-1999:	99DE-1042125.

PI Pompejus M., Kroegeer B., Schroeder H., Zelder O., Habernauer G.
XX
DR WP1: 2001-061975/07.
DR N-PSDB; AAF71518.
XX

PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
PT

PS Claim 20; Page 587-588; 1246pp; English
v v

CC AAF7136 to AAF7150 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a prolineogenic
CC or nonprolineogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a
CC cofactor, a polyelectrolyte, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (1.e. ATP, NADPH).

Sequence 295 AA;

Query Match	24.98;	Score 349;	DB 22;	Length 295;
Best Local Similarity	34.58;	Pred. No. 1.1e-26;		
Matches	86;	Conservative	45;	Mismatches 114;
			Indels	4;
			Gaps	2

OY	30	LAKKALVASTDGIIGFIRARRLADOGAAHYVS--SKOONDOQAVATLOEGELSVTGTV	87
Db	48	Ikyrkaliqsgsigaaavaiaayaregadaiaiyipeegdadadrvigaletygfakafsp	107
OY	88	CHVKAEDEBERIVAAVLHGGIILIVSMAANVPFGESIMDTVEEWMDKTLDINKAPAL	147
Db	106	gdrlpeyrslvqetvalagldilvynasrqwavgjltelvdnfdqlvnylygsfr	167
OY	148	MTKAAYPEMEKKGGGSVIYSIAAFSPSCFSYINVKSTALLGLTKTLATELPARNIRV	207
Db	168	vtkaaipnl-kpgssilfftsiqaypsettildyamtkaalnnlstkjassilfgdqjrv	225
OY	208	NCLAGGLKTSESRMLMWDKEKSEMETLRIIRLRLEPEDCAGIYSLFCSDASAIYTGET	267

db

```
33 KVALVTFASTDGI GFALIRRLAQDGAHVVS-SRKQNVDAVATLÖGGLSVTGVCHWG 91  
| |||| :: ||| :|| :||::: | : : : : | : : : :  
5 ksalvtgasrigrsialqlaeegynvavnyagskeaaaveelkaigvdvsfalqanva 64
```

biological agents affecting fatty acid biosynthesis, the method is used for screening biological agents affecting fatty acid biosynthesis. Agonists and antagonists of fab (fatty acid biosynthesis) are used to inhibit, prevent or treat diseases such as infections of the upper respiratory tract (e.g. otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (e.g. empyema, lung abscess), cardiac (e.g. infective endocarditis), gastrointestinal (e.g. secretory diarrhoea, pleuritic abscess, retroperitoneal abscess), central

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:16:31 ; Search time 16.41 Seconds

(without alignments)
1290.465 Million cell updates/sec

Title: US-09-866-034-2

Perfect score: 1399

Sequence: 1 MHKAGLGLICARAAVSVRNA.....DASYITGETVIVGGTIPSRL 278

Scoring table:

BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR.68: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	824.5	58.9	280	2	S6665	nuclear protein H
2	636	45.5	257	2	T32002	hypothetical prot
3	614	43.9	260	2	T22676	hypothetical prot
4	588.5	42.1	254	2	E85069	hypothetical prot
5	459.5	32.8	260	2	E70881	probable fabG5 pro
6	409	29.2	255	2	D83416	probable short-cha
7	399	28.5	246	2	A69621	3-oxoacyl-(acyl-ca
8	396.5	28.3	246	2	H72219	3-oxoacyl-(acyl-ca
9	388	27.7	246	2	C83961	3-oxoacyl-(acyl-ca
10	385	27.5	248	2	H70447	3-oxoacyl-(acyl-ca
11	381.5	27.3	282	2	C71204	probable dehydroge
12	381	27.2	255	2	H84288	3-oxoacyl-(acyl-ca
13	374	26.7	243	2	F86721	hypothetical prote
14	373.5	26.7	251	2	G73389	oxidoreductase, sh
15	371	26.5	263	2	C75217	probable short-cha
16	370	26.4	249	2	B83767	glucose 1-dehydrog
17	369.5	26.4	285	2	D85885	probable oxidoredu
18	368.5	26.3	285	1	A65017	probable 3-oxoacyl
19	368.5	26.3	320	2	S32450	3-oxoacyl-(acyl-ca
20	366	26.2	261	2	S35196	hypothetical prote
21	365.5	26.1	248	2	F69688	glucose 1-dehydrog
22	364.5	26.1	260	2	B48674	troponine reductas
23	364	26.0	254	2	G53533	troponine reductas
24	360.5	25.8	261	2	S54815	uridine polyket
25	360	25.7	272	1	S05397	gramicidin polyket
26	358.5	25.6	256	2	A70637	hypothetical prote
27	357	25.5	248	2	F81971	probable 3-oxoacyl
28	357	25.5	248	2	E81026	3-oxoacyl-(acyl-ca
29	357	25.5	273	1	A48674	troponine reductas

30	354	25.3	241	2	F71636	3-oxoacyl reductas
31	354	25.3	278	2	T10877	y41A protein - Rhl
32	353	25.2	262	2	B84695	probable troponine
33	349	24.9	256	2	E72427	oxidoreductase, sh
34	349	24.9	258	2	G69755	glucose 1-dehydrog
35	347.5	24.8	253	2	E83207	probable short-cha
36	345.5	24.7	268	2	B84693	probable troponine
37	345	24.7	246	1	DEKCN	acetoacetyl-CoA re
38	345	24.7	248	2	D70707	hypothetical prote
39	344.5	24.6	255	2	G82644	2,5-dichloro-2,5-c
40	343	24.5	248	2	F82128	3-oxoacyl-(acyl-ca
41	342	24.4	247	2	S71280	3-oxoacyl-(acyl-ca
42	341.5	24.4	246	2	H84136	3-oxoacyl-(acyl-ca
43	341	24.4	246	2	E81695	probable troponine
44	341	24.4	268	2	E84695	2-deoxy-D-gluconat
45	341	24.4	281	2	F69400	

ALIGNMENTS

```

RESULT 1
S6665
nuclear protein Hep27 - human
N:Alternate names: protein D
C:Species: Homo sapiens (man)
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S6665; S66614; S14094; S78122
R:Gabrielli, F.; Donadel, G.; Bensl, G.; Heguy, A.; Melli, M.
A:Title: A nuclear protein, synthesized in growth-arrested human hepatoblastoma cell.
A:Reference number: S66614; MUID:96035681
A:Accession: S6665
A:Molecule type: mRNA
A:Residues: 1-280 <GAW>
A:Cross-references: EMBL:U31875; NID:G1079565; PID:AA82048.1; PID:G1079566
A>Note: Met-23 is the probable initiator
A:Accession: S66614
A:Molecule type: protein
A:Residues: 24-41 <GAW>
R:Donadel, G.; Gabrielli, C.; Frank, R.; Gabrielli, F.
Eur. J. Biochem. 195, 723-729, 1991
A:Title: Identification of a novel nuclear protein synthesized in growth-arrested hu
A:Reference number: S14094; MUID:91153312
A:Accession: S14094
A:Molecule type: protein
A:Residues: 80-83, 'L', 85, 87-88, 'A', 90-91, 141-146, 148-162, 198-205 <DON>
A:Accession: S78122
A:Molecule type: protein
A:Residues: 'V', 81-84, 'G', 86-88 <DON>
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: nucleus
F:37-216/Domain: short-chain alcohol dehydrogenase homology <SABH>

Query Match 58.9% Score 824.5; DB 2; Length 280;
Best Local Similarity 61.6% Pred. No. 3.4e-58;
Matches 162; Conservative 47; Mismatches 51; Indels 3; Gaps 2;

OY 10 CARAAVSVRNAAASGMRTRDPLANKVALVYASTDGIQRIARRLAQDGAHVVSRRKQOV 69
DB 16 CARL--SVRNASTSIDRKGVLANVAVVGTSGIGFAIARLRADGAHVVISRRKQOV 73
OY 70 DOAAVATLQGEGLSVTGVCHVCHGKAEDREBLVATVAKLHGGLDILVNAANVPFGSTMDV 129
DB 74 DRMAKIQGEGSLVAGTVCHVCHGKAEDREQLAKALEHCGGVDFVCSAGVNPVLTGTLGT 133
OY 130 TEVYMDKTLIDINKAPALMKRAVPEMEKRGCGSVTVSSIAFSPSPGSPVNSKTAL 189
DB 134 SEQIWDKTLISVNVKSPALLLSQLPYMENR--RGAVILVSSIAVNPVVALGVTVNSKTAL 192
OY 190 LGILKTLATLAPRNTIVNCLAPGLKLTSTFSRLMDKEESNKETLIRIRLGEPEPCA 249
DB 190 LGILKTLATLAPRNTIVNCLAPGLKLTSTFSRLMDKEESNKETLIRIRLGEPEPCA 249

```

```

Db      193 LGTPTLTALELPKADIRVNCVPGLIKTDSENVFNGNESLMKNREKHOLRIQGESEDCA 252
OY      250 GIVSFILCEDASTITGETTVVGCG 272
          ||||| | | | | : : : |
Db      253 GIVSFLCSPDASVNGENIAVAG 275

RESULT 2
T32002
hypothetical protein F36H9.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
R:Accession: T32002
R:Daute, M.; Kramer, J.; Twyman, B.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid F36H9.
A:Reference number: Z21110
A:Accession: T32002
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-257 <DAN>
A:Cross-references: EMBL:AF016668; PIDN:AAB66092.1; GSPDB:GN00020; CESP:F36H9.3
A:Experimental source: strain Bristol N2; clone F36H9
A:Genetics:
A:Gene: CESP:F36H9.3
A:Map position: 2
A:Introns: 21/3; 49/3; 192/1
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match           45.5%; Score 636; DB 2; Length 257;
Best Local Similarity 50.6%; Pred. No. 2.8e-43;
Matches 126; Conservative 49; Mismatches 74; Indels 0; Gaps 0;

OY      30 LANKVALVTASDGIGFAIRRLAODGAHVYVSRRKOONDQAVATLGEGSLVTGTCH 89
          ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      9 LTDRALVAIYTSRKGGIFAKIQLAGASVVYGSRKKENVDEAAALARLENIDAHOTTAA 68
          .....: | | | | | | | | | | | | | | | | | | | | | |
OY      90 VGKAEDRELVAATAVKELHGIDIIYSNAAVNPFGSGIMDTVEEWQKTIDIYNKAPALMT 149
          : : : | : : : | | | | | | | | | | | | | | | | | | |
Db      69 VGNKSDDRKLTLDFLTRFTKLDILLISNAAVNPHTGDLKMTDSQMOKLLIDLNKSAFELT 128
          ..: | | | | | | | | | | | | | | | | | | | | | | |
OY      150 KAVNEEMKRGGGSVIYSSSTAAPSPPGFYVNSKTALLGLTKLTALELRINRVNC 209
          | : : | : | | | | | | | | | | : | : | | | | | | | |
Db      129 KEAVNHLEASSRGNAVVFSSVAGYSPMNEGAYSVMKTTLTGLSKSIALLARNRIVNS 188
          . | : | : | | | | | | | | | | : | : | | | | | | | |
OY      210 LAPGLIKTFESRMMLMPDKEREESMEKEPTRLIRIGEPEPDACGIYSLCESDASTITGETTV 269
          : | | | | | | | | | | | | | | | | | | | | | | | | |
Db      189 IAPGITDIQDFEQLVFSDESEKQKWLSQAIDRRFDPECAEAVALFYSDIASITSGETTG 248
          : | | | | | | | | | | | | | | | | | | | | | | | | |
OY      270 VGGGTPTSRL 278
          : | | : | | |
Db      249 INGMWHARI 257

RESULT 3
T32676
hypothetical protein F54F3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
R:Accession: T32676; T24959
R:Percy, C.; Lloyd, C.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19598
A:Accession: T32676
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-260 <WLL>
A:Cross-references: EMBL:Z79696; PIDN:CAB01974.1; GSPDB:GN00023; CESP:F54F3.4
A:Experimental source: clone F54F3
R:Lloyd, C.
submitted to the EMBL Data Library, November 1996
Reference number: Z19960
```

A:Accession: T24959
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <MI2>
A:Cross-references: EMBL:281592; PIDN:CAB04734.1; GSPDB:GN00023; CESP:F54F3.4
A:Experimental source: clone TI601
C:Genetics:
A:Gene: CESP:F54F3.4
A:Map position: 5
A:Introns: 21/2; 81/3; 201/3
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 43.9%; Score 614; DB 2; Length 260;
Best Local Similarity 48.0%; Pred. NO. 1.6e-41;
Matches 120; Conservative 56; Mismatches 70; Indels 4; Gaps 2;

QY 33 KAAIVTASNDGIGGFATARLADGAAHYVYSSKKQONDOAVATLOGEGLS-VTGTVCHVG 91
DB 11 KAAIVTAAATKGIIGLAEHLDEGASVYVGSNNQNDVEALIEYLNKKGITKVAAGIAGHIA 70
QY 92 KAEDERLVAATVAKLHGGLDILVNAAVPFGSLMDYTEEYWDKTLIDINAKAPALMKR 151
DB 71 STDDQKIVLDFLQKFGKINILVNHGIVPATGHLLEVSQVWDKLFEEVNAKAGQMKRL 130
QY 152 VPEMEKRGGSVVIVSSIAAFSPGSPYVASTALLGLTKTIAELAPRNIRVNCIA 211
DB 131 VHPHIAKEGGAIIFNASYSAYKSPPGIAAYGVTKTLVLGLRALAMLANDINRVNGIA 190
QY 212 PGLITSPSRMVL---MKEKEESMKELRLIRRLCEPEDECAQISFLCSEASVITGETV 268
DB 191 PGLVITKKSQVILMDGGEAEKELTDIOETALGELVDPDDCAGTAVAYLASDDSSSYITGEMI 250
QY 269 VVGCGTSPRL 278
DB 251 ILAGVQARL 260

RESULT 4
885069
hypothetical protein AT4g05530 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Dates: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: E85069
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: E85069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:Cross-references: GB:NC_001268; NID:g7267313; PIDN:CAB81095.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g05530
A:Map position: 4

Query Match 42.1%; Score 588.5; DB 2; Length 254;
Best Local Similarity 50.6%; Pred. NO. 1.7e-39;
Matches 126; Conservative 35; Mismatches 85; Indels 3; Gaps 2;

QY 30 LANNVALTASTDGIQFAIARLADGAAHYVYSSKKQONDOAVATLOGEGLSVTGVCH 89
DB 9 LEGKVAIVTASTGIGITGIFTERFGEAGSVYSSRKQANDENAVAKIKSGIDAYGIVCH 68
QY 90 VGRAEDDERLYATVAKLHGGLDILVNAAVPFGSLMDYTEEYWDKTLIDINAKAPALMT 149
DB 69 VSNQHRNLIVTEKTVQKGIIDIVVCAANPSTDPILSKSEAVLDKLTMEINVKSSITLL 128
QY 150 KAAVPEMEKRGGSVVIVSSIAAFSPGSPYVASTALLGLTKTIAELAPRNIRVNC 209
DB 129 QDMAVPLEK--GSSVITFTTLAGFSPOGAAAMGVTTALLGLTKLAAMAP--DTRVNA 185

A/Accession: D83416

R; Cronan, J. E.; Mor

R; Cronan, J.E.; Morbidoni, H.R.; de Mendoza, D.
References: DDBJ: D04110; NID: g1389348; P1BDN: BAA109/4.1; PID: g1237012

R; Cronan, J.E.; Morbidoni, H.R.; de Mendoza, D.
References: DDBJ: D04110; NID: g1389348; P1BDN: BAA109/4.1; PID: g1237012

RESULT 11
C71204
probable dehydrogenase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: C71204
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchikawa, N. Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon, Pyrococcus horikoshii
A:Reference number: A71000; M0ID:98344137
A:Accession: C71204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-282 <KAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BA031026.1; PID:g3258343
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:gene: PH1901
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F:28-207/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match	27.38;	Score 381.5;	DB 2;	Length 282;
Best Local Similarity	37.7%;	Pred. No. 5.2e-23;		
Matches 98;	Conservative 49;	Mismatches 100;	Indels 13;	Gaps 6;

Query Match	27.28;	Score 381;	DB 2;	Length 255;
Best Local Similarity	35.38;	Pred. NO. 5e-23;		
Matches 89;	Conservative 47;	Mismatches 106;	Indels 10;	Gaps 4

[illegible]

RESULT	13
F86721	
hypothetical protein fabG1 (imported) - Lactococcus lactis subsp. lactis (strain IL14)	

Db 6 KVLITGSGGICKAAVMEARGAKVAINDISEEKETVELIKSMGGEAIFIGDV-- 63

Search completed: May 8, 2002, 08:24:40
Job time: 400 sec

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
E:9-188/Domain: short-chain alcohol dehydrogenase homology <SAD>

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Colton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL genome sequence of *Thermotoga maritima*.";
 CC Nature 399:323-329(1999).
 CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001811; AAD36790.1; -
 DR TIGR: TM1724; -
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRDH.
 DR PRINTS: PR01167; INSADHPFAMILY.
 KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 154 154 BY SIMILARITY.
 SQ SEQUENCE 246 AA; 26401 MW; 8C08904D28099142 CRC64;

Query Match 28.3%; Score 396.5; DB 1; Length 246;
 Best Local Similarity 36.5%; Pred. No. 1.4e-22;
 Matches 89; Conservative 54; Mismatches 98; Indels 3; Gaps 3;

QY 30 LANKVALVTASTDGIQAIARRLAODGAHVYSSRKQONDQAVATLOGESLSTYGTCH 89
 DB 3 LEKVCLITGAAGIGATLTLRQEGATVIADISKENDLSYKKEGJLPGKVDYVLN 62
 QY 90 VCAEDRERLVATAVKLHGIDILVSNAAVNPFGSIMDYTEEYMDKTLIDINVAAPALMT 149
 DB 63 VTDRDQIKVEYERVKYGRIDVLVNNAGITR-DALLVRKKEEDMDAVINVLKGVENVT 121
 QY 150 KAVPEPEKRGSGSVYVSSIAFSPGSPYNSKTLALLGLTKTLAIELAPRINIVNC 209
 DB 122 QMVPVPMIKQKNSIVNVSSVGIYGNPOTNYAASKAGVIGMTWAKELAGRNIRVNA 181
 QY 210 LABGLIKTSFRLMMDKEESMKETLIRIRLGEPEDCAGISYFLCSLSEASTITGTVV 269
 DB 182 VAGFLETPTKTL-PEKARETALSR-IPLGKRGKPEVAQVILLFLASDESSYVTGQVIG 239
 QY 270 VGGG 273
 DB 240 IDGG 243

RESULT 4
 FABG_AQUAE STANDARD; PRT; 248 AA.
 AC 067610;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
 DE ACYL CARRIER PROTEIN REDUCTASE).
 GN FABG OR AQ_1716.

OS Aquifex aeolicus.
 OC Bacteri; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxId=63363;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5.
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young M.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Anuj M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus".
 RL Nature 392:353-358(1998).
 CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000752; AAC07575.1; -
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRDH.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 12 36 NADP (BY SIMILARITY).
 FT ACT_SITE 157 157 BY SIMILARITY.
 SQ SEQUENCE 248 AA; 26867 MW; 5CFD9B9ADB83F2C5 CRC64;

Query Match 27.5%; Score 385; DB 1; Length 248;
 Best Local Similarity 37.8%; Pred. No. 1e-21;
 Matches 95; Conservative 49; Mismatches 91; Indels 16; Gaps 6;

QY 30 LANKVALVTASTDGIQAIARRLAODGAHVYSSRKQONDQAVATLOGESLSTYGT 86
 DB 5 LQGVSLVTGSGTIGRAIAEKLASGTVIITGSERAKAVAEIANKY-GYKAGV 62
 QY 87 VCHGKAEDRERLVATAVKLHGIDILVSNAAV--NPF--GSLMDYTEEYMDKTLIDIN 142
 DB 63 EMLISEESIKAPFEELYINLVGDIDILVNNAGITRDLFLRMSILD-----WEEVLKVN 117
 QY 143 KAPALMTKAVPEPEKRGSGSVYVSSIAFSPGSPYNSKTLALLGLTKTLAIELAP 202
 DB 118 TGFPLVQLNSLRKIKRMRGRIVNISSVGFNGVGVNSTKAGLIGTKSLAKELAP 177
 QY 203 RNRIVNCLAPGLIKTSFRLMMDKEESMKETLIRIRLGEPEDCAGISYFLCSLSE 262
 DB 178 RNVLVNNAVADPFIETDITAVL--SEIKQYKKGQIPILGRGSPPEVANNVLLFSL 235
 QY 263 ITGETVYVGGG 273
 DB 236 ITGETVYVNGG 246

RESULT 5
 TRN2_HYONI STANDARD; PRT; 260 AA.
 AC P50164;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

Query Match	27.1%	Score 379.5;	DB 1;	Length 260;
Best Local Similarity	35.4%	Pred. No. 2.8e-21;		
Matches 90; Conservative	49;	Mismatches 110;	Indels 5;	Gaps 3;

[illegible]

Query March	26.33;	Score 368.5;	DB 1;	Length 320;
Best Local Similarity	33.58;	Pred. No. 2.3e-20;		
Matches 91;	Conservative 56;	Mismatches 102;	Indels 23;	Gaps 6;
QY	19	MASSGMFRDRLANKVA-----LVTASDGIQFIARRLRADQAHVVY-----S 62		
	:	:		

ID	TRN2	DATST	STANDARD	PRT	260 AA
Db	52	IATSGINAQVATLAKKVSAGAGQGVESVYVLTGASRGTCAINALSLCKACQKVLVYNARS			111
Qy	63	SRKOONDOQAVATLQGEGLSVTGCVHGAEDRERLVATAVKHLHGIDILVSNAAVNPF			122
Db	112	SKKEEVEVSKETIEAFGGQALTFEGC---	DYSKDEEDVEMIKTAVAMQGVTVDLVNNAGITR-		167
Qy	123	FGSIMDTEEMVMDKTLIDINKAPRLMKRAYPEMEKRGKGSVVIVSIAAFSPSPGFSPY			182
Db	168	DGLIMRRKKQKQWQOVEDLNLITGVFLCQAQAKIMKKKKRIIINIASVGLVJNAGQANY			227
Qy	183	NVSKTALLGCTTKTLATIELAPRINRVNCLAPKLTSPSRMLMMDKKEEBSMKETLRRL			242
Db	228	SAKAGVIGTGTAVKARYASRNINVAAGCFISSDMTSKLGODINK--	ILEFTIPLGRY		285
Qy	243	GEPEDCAGIVSFLCSEDA-SYITGETVWVGQ			273
Db	286	GOPEEVAGIVEFLAIPASSVYTGQVETIDGG			317
RESULT	8				
ID	TRN2	DATST	STANDARD	PRT	260 AA
AC	50163				
DT	01-OCT-1996	(Rel. 34, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	TROPINONE REDUCTASE-II (EC 1.1.1.236)	(Tr-II).			
GN	TR2				
OS	Datura stramonium (Jimsonweed) (Common thornapple).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;				
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;				
OC	Asteridae: eusterids I; Solanales; Solanaceae; Datura.				
OX	NCBI_TaxID=4076;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=ROOT;				
RX	MEDLINE=94022421; PubMed=8415746;				
RA	Nakajima K., Hashimoto T., Yamada Y.;				
RT	"Two tropinone reductases with different stereospecificities are				
RL	short-chain dehydrogenases evolved from a common ancestor.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:9591-9595(1993).				
RN	[2]				
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).				
RX	MEDLINE=98226735; PubMed=9560196;				
RA	Nakajima K., Yamashita A., Akama H., Nakatsu T., Kato H.,				
RT	Hashimoto T., Oda J., Yamada Y.;				
RL	"Crystal structures of two tropinone reductases: different reaction				
RL	stereospecificities in the same protein fold."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:4876-4881(1998).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RX	MEDLINE=99316165; PubMed=10387002;				
RA	Yamashita A., Kato H., Makatsuki S., Tomizaki T., Nakatsu T.,				
RA	Nakajima K., Hashimoto T., Yamada Y., Oda J.;				
RT	"Structure of tropinone reductase-II complexed with NADP+ and				
RT	pseudotropine at 1.9 A resolution: implication for stereospecific				
RL	substrate binding and catalysis."				
RL	Biochemistry 38:7630-7637(1999).				
CC	-1- FUNCTION: CATALYZES THE STEREOSPECIFIC REDUCTION OF TROPINONE TO				
CC	PSEUDOTROPINE.				
CC	-1- CATALYTIC ACTIVITY: PSEUDOTROPINE + NADP(+) = TROPINONE + NADPH.				
CC	-1- PATHWAY: BIOSYNTHETIC PATHWAY OF TROPANE ALKALOIDS.				
CC	-1- SUBUNIT: HOMODIMER.				
CC	-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES				
CC	(SDR) FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				

```

CC EMBL: L20474; AAA33282.1;
DR PDB: 2AE2; 18-NOV-98.
DR PDB: 2AE2; 02-FEB-99.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDHRDH.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP: 3D-structure.
FT NP_BIND 13 37 NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 260 AA; 28311 MW; 2DBF4963B2CACA303 CRC64;

```

```

Query Match
Best Local Similarity 34.6%; Score 364.5; DB 1; Length 260;
Matches 88; Conservative 50; Mismatches 111; Indels 5; Gaps 3;

```

```

OY 24 MTRRDEPLANKVALVASTDGIQFAIRRLAQDGAHVYSSRKQONVDAVATLOGEGSLV 83
DB 1 MAGRWNLLECTALVTGSGSIGIYVIELASLGASVYTCGRNCKELDTQNRSGKFV 60
OY 84 TGVCHVGAEDRELRVAT-AVKLHGIDILVSNAAVNPFFGSDIMDTEWMDTKIDINV 142
DB 61 EASVCLSSRSEKOEIMNTVANHFGKLNILVNNAGI-VYKAKYTVEDYSLINSIN 119
OY 143 KAPALMTKAVPEMEKRGGSVIVASSIAFSPSPSPINVKTLGLTKTLAIELAP 202
DB 120 EAAVHLSTVAHPFLKASERGNVVFISVSGALAVPEAVGATKGMDOLTCLAEEMAK 179
OY 203 RIRIRVNCAPLKLKTSFSLMMDKEKESMKETL--RIRRLGPEDDAGTIFLCS 259
DB 180 DNIIRVNGVGPVYIATSLVEMTIQDPEQKLNKILRICALRNKGEKELAAVAFICFPA 239
OY 260 ASYITGETVYVGGG 273
DB 240 ASYVTGQITLYVDGG 253

```

```

RESULT 9
UCPA_ECOLI STANDARD; PRT; 263 AA.
AC P37440; P77442; P76963; P77140;
DT 01-OCT-1994 (Rel. 30, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OXIDOREDUCTASE UCPA (EC 1.-.-.-).
GN UCPA OR B2426.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Petina N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,

```

```

RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50,0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97385354; PubMed=9241368;
RA Sirko A., Weglenska A., Hryniewicz M.M., Hulanicka D.M.;
RT "Characterization of the Escherichia coli gene encoding a new member
RT of the short-chain dehydrogenase/reductase (SDR) family."
RL Acta Biochim. Pol. 44:153-157(1997).
RN [4]
RP SEQUENCE OF 180-263 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90264335; PubMed=2188959;
RA Hryniewicz M.M., Sirko A., Palucha A., Boeck A., Hulanicka D.M.;
RT "Sulfate and thiosulfate transport in Escherichia coli K-12:
RT identification of a gene encoding a novel protein involved in
RT thiosulfate binding."
RL J. Bacteriol. 172:3358-3366(1990).
RN [5]
RP IDENTIFICATION.
RX MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome."
RL Nucleic Acids Res. 22:4756-4767(1994).
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000330; AAC75479.1; ALT_INIT.
CC EMBL: D90872; BAA16309.1; ALT_INIT.
CC EMBL: D90871; BAA16300.1;
CC EMBL: X99908; CAA68181.1;
CC EMBL: M32101; -, NOT_ANNOTATED_CDS.
CC HSSP: P14061; IFDV.
CC Ecogene; EG12133; ucpa.
CC InterPro: IPR002198; ADH_short.
CC InterPro: IPR002347; Adh_short_C2.
CC Pfam: PF00106; adh_short_1.
CC Pfam: PF00678; adh_short_C2; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; FALSE_NEG.
KW Oxidoreductase; Complete proteome.
FT CONFLICT 153 153 T->Q (IN REF. 3)
FT SEQUENCE 263 AA; 27850 MW; 2E201713357FDF41 CRC64;

```

```

Query Match
Best Local Similarity 34.1%; Score 362.5; DB 1; Length 263;
Matches 86; Conservative 53; Mismatches 104; Indels 9; Gaps 4;
OY 30 IANKVALVASTDGIQFAIRRLAQDGAHVYSSRKQONVDAVATLOGEGSLVGTGCH 89
DB 4 LTGKTAALTALGALOGIGEGIAFTFARGANLLIDISPE-IEKADLCGRGRCRTAVAD 62
OY 90 VKAADRELVATFAVKLHGIDILVSNAAVNPFFGSDIMDTEWMDTKIDINVKAPALMT 149
DB 63 VQDPASVAALIRAKREKGRIDILVNNAGVCR-LGSEFLDMSDDDRDFHIDINIKGVNVT 121
OY 150 KAVVPEMEKRGGSVIVASSIAA-FSPSPGFSPYNNVKALLGLTKTLAIELAPRIRIN 208

```

```

DB 122 KALPELMARKDRIWMSVYTGDMVADPGETAYALTKAIVGLTSLAVEYASQSGIRVN 181
      ||:|||| | | :|||: ||: | :||:||||:| | ||||
QY 209 CIAPGIIKTSFSLMWNDEKEE-----SMKETLRIRRGEPEDCAGIYSLCSEBASY 262
      | | :| | :| | :| | :| | :| | :| | :| | :| |
DB 182 AICPGIVRTPMASIRASQNPEDPESVLTETMAKAIPIRLADPLEVGEALFASDESSY 241
      || | | | | | | | | | | | | | | | | | |
QY 263 TGETVVGCGT 274
      || | | | | | | | | | | | | | | | | | |
DB 242 LIGTONVIDGGS 253
      || | | | | | | | | | | | | | | | | | |

RESULT 10
DHKL_STRVN STANDARD: PRT: 272 AA.
AC P16542:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE GRANATICIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE 1
   (EC 1.3.1.-) (ORF5).
OS Streptomyces violaceoruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1935;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TU22;
RX MEDLINE=90060034; PubMed=2583128;
RA Sherman D.H., Maipartida F., Bibb M.J., Kieser H.M., Bibb M.J.,
RA Hopwood D.A.;
RT "Structure and deduced function of the granaticin-producing
RT polyketide synthase gene cluster of Streptomyces violaceoruber
RT Tu22."
RL EMO J. 8:2717-2725(1989).
CC -I- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC GRANATICIN.
CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X16300; CA34368.1; -
DR EMBL: X16144; CA34263.1; -
DR PIR: S05397; S05397.
DR HSSP: P03617; 1IFK.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDRHD.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT NP_BIND 21 45 NAD (BY SIMILARITY).
FT ACT_SITE 168 168 BY SIMILARITY.
SO SEQUENCE 272 AA; 28393 MW; 0DE5415A19FB218 CRC64;

```

```

DB 76 ADPAQIRAYAAARQRTGYDILVNNAG-RSGGATREIDELMDVITTNLTSVPLMTK 134
      ||| | | | | | | | | | | | | | | | | |
QY 151 AYVPE-----MEKRG-----GGSVIVSSIAFSPSPSPYNNKVTALLGLTFLAI 198
      | | :| | | | | | | | | | | | | | | | | |
DB 135 EVLNAGCMALKKRRIININASTGKGQGVHAY-----PYASKHGVTGTLALGL 184
      || | | | | | | | | | | | | | | | | | |
QY 199 ELAPRNIRVNCIAPGLTKTSFSR-----LWMDKKESEKETLRI--RIUGPEDCA 249
      ||| | | | | | | | | | | | | | | | | |
DB 185 ELARTGITVANVCGFVETMAERVRHRYAGIMOVSEETFRITNVPILGRVYETREVA 244
      | | | | | | | | | | | | | | | | | | |
QY 250 GIVSFLCEDASYITGETVVGCG 273
      | | | | | | | | | | | | | | | | | | |
DB 245 AMVEYLVAADDAVAATQAALVNCGG 268
      | | | | | | | | | | | | | | | | | | |

RESULT 11
TRNL_DATST STANDARD: PRT: 273 AA.
AC P50162:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TROPINONE REDUCTASE-I (EC 1.1.1.206) (TR-I) (TROPINE DEHYDROGENASE).
TR1.
OS Datura stramonium (Jimsonweed) (Common thornapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Datura.
OX NCBI_TaxID=4076;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Root;
RX MEDLINE=94022421; PubMed=8415746;
RA Nakajima K., Hashimoto T., Yamada Y.;
RT "Two tropinone reductases with different stereospecificities are
RT short-chain dehydrogenases evolved from a common ancestor."
RL Proc. Natl. Acad. Sci. U.S.A. 90:9591-9595(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=98226735; PubMed=9560196;
RA Nakajima K., Yamashita A., Akama H., Nakatsu T., Kato H.,
RA Hashimoto T., Oda J., Yamada Y.;
RT "Crystal structures of two tropinone reductases: different reaction
RT stereospecificities in the same protein fold."
RL Proc. Natl. Acad. Sci. U.S.A. 95:4876-4881(1998).
CC -I- FUNCTION: CATALYZES THE STEREOSPECIFIC REDUCTION OF TROPINONE TO
CC TROPINE.
CC -I- CATALYTIC ACTIVITY: TROPINE + NADP(+) = TROPINONE + NADPH.
CC -I- PATHWAY: BIOSYNTHETIC PATHWAY OF TROPANE ALKALOIDS.
CC -I- SUBUNIT: HOMODIMER.
CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L20473; AAA33281.1; -
DR PDB: 1AE1; 18-NOV-98.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00081; GDRHD.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; 3D-structure.
FT NP_BIND 25 49 NADP (BY SIMILARITY).
FT ACT_SITE 171 171 BY SIMILARITY.

```


SO SEQUENCE 273 AA: 29617 MW: 39A523EF04EAB1F1 CRC64:

Query Match 25.5%; Score 357; DB 1; Length 273;
Best Local Similarity 33.6%; Pred No. 1,4e-19;

Matches 89; Conservative 58; Mismatches 110; Indels 8; Gaps 4;

QY 15 NSVRASSGNTRRDPLANKVALVASTDGIQFIARRLDGDAHVVSSRKQONVQAV 74
D 4 SKVSMNCCNNEGHMSLKGTTALVTGSGKIGVAIVELAGLGRVYTCSSNNEELDECLE 63
QY 75 TLQDEGLSTGYCHNGKAKEDRRLVATVAKLHG--IDLIVSAAVNPFGSIMDYEEV 133
D 64 IMREKGLNVEGVCYDLISRTEDKIMQTAHVFDGKRLNLTNNAGV-VIHKAKDFEKD 122
QY 134 WDKTLINVKAPALMTKAVVPEMEKRGGSVYVSSIAFSPSPGSPYVNTALLGLT 193
D 123 YNIMKTNEAAYHLSQIAVPLIKASQNVFLSSIAAGSALPVSLSASGAINOMT 182
QY 194 KTLAIFLAPRINRVNCLABG-----LKTSPSRMLMDREKESMKETLRIRLGEPEDC 248
D 193 KSLACEMANDNIRVNSVARGVILFPLEVTAIKKNPHQKEIDNFIYKT-PMGRAGKPEY 241
QY 249 AGVSVLCESDASYTGTETVVVGGG 273
D 242 SALIAFLCPAASYITGQIIMADGG 266

RESULT 12

FABG_RICPR STANDARD: PRT: 241 AA.

AC P50941;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
ACYL CARRIER PROTEIN REDUCTASE).
GN FABG OR RP762.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiase; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sierberitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE OF 1-41 FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=94179117; PubMed=8132476;
RA Dunkin S.M., Winkler H.H., Wood D.O.;
RT "Isolation and characterization of the Rickettsia prowazekii recA
RT gene.";
RL J. Bacteriol. 176:1777-1781(1994).
RN [3]
RP IDENTIFICATION.
RX MEDLINE=96259307; PubMed=8662004;
RA Andersson S.G.E., Sharp P.M.;
RT "Codon usage and base composition in Rickettsia prowazekii.";
RL J. Mol. Evol. 42:525-536(1996).
CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: A1235273; CAAL1590.1;
DR EMBL: U01959; -, NOT_ANNOTATED_CD5.
DR HSSP: Q12634; 1YBY.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY;
DR PRINTS: PR00081; GDRDH.
DR PRINTS: PR01167; INSADHFAMILY.
DR PROSITE: P50061; ADH_SHORT_1.
KM Fatty acid biosynthesis; Oxidoreductase; NADP: Complete proteome.
FT NE_BIND 11 NADP (BY SIMILARITY).
FT ACT_SITE 148 148 BY SIMILARITY.
SO SEQUENCE 241 AA: 25759 MW: E43B8711545B8295 CRC64:

Query Match 25.3%; Score 354; DB 1; Length 241;
Best Local Similarity 33.6%; Pred. No. 1.9e-19;
Matches 82; Conservative 61; Mismatches 91; Indels 10; Gaps 4;

QY 30 LANKVALVASTDGIQFIARRLDGDAHVVSSRKQONVQAVATLQDEGLSTGYCH 89
D 4 LTKTSLITGASSGIGSATAIRLHLKSGKYIISGSNEEKLSIGNALKD---NYTTEVCN 60
QY 90 VCGAEDRERLVAATVAVLHGIDILVNSAAVNPFGSIMDYEEVMDKTDINVKAPALMT 149
D 61 LANKECNSMLISKT-----SNLDILVCNAGITSDPLAIR-MKDDPEKVIDINKAFILM 115
QY 150 KAVVPEMEKRGGSVYVSSIAFSPSPGSPYVNTALLGLTKTLAIFLAPRINRVN 209
D 116 REAIKRMIDKRGRIINISSIVGIAQNPQANVCASKAGLIGMTKLSYEVAATRGITVNA 175
QY 210 LAPGLKTSFSRMLMDREKESMKETLRIRLGEPEDCAGIYSELCSDASYTGTETV 269
D 176 VAPGFIKSDMTDKL-NEKOREAIVOKITPLGTYGIPEDVAVAFAVANNASNYITGOTLH 233
QY 270 VGGG 273
D 234 VNGG 237

RESULT 13

YALA_RHISN STANDARD: PRT: 278 AA.

AC P55541;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PURATIVE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE YALA (EC 1.-.-.-).
GN YALA.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.

DE STRESS PROTEIN 74) (GSP74).
GN YCDE.
OS Bacillus subtilis.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the *lin*comycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the *lin*-2 mutation";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RN [2].
RP SEQUENCE OF 1-13.
RC STRAIN=IS58;
RX MEDLINE=97443988; Pubmed=9298659;
RA Anselmann H., Bernhard J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -1- CATALYTIC ACTIVITY: BETA-D-GLUCOSE + NAD(P)(+) = D-GLUCONO-DELTA-
CC LACTONE + NAD(P)H.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
CC LIMITATION AND OXYGEN LIMITATION.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: AB000617; BAA2244.1; -
DR EMBL: Z99105; CAB12077.1; -
DR SUDILast; BG12761; YCDE.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short.C2.
DR Pfam: PF00106; adh_short.1.
DR Pfam: PF00678; adh_short.C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDRDH.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 11 35 NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 258 AA; 27776 MW; 77842DC45D96C26 CRC64;

Query Match	24.98;	Score 349;	DB 1;	length 258;
Best Local Similarity	33.98;	Pred. No. 4.9e-19;		
Matches 87; Conservative	53;	Mismatches 105;	Indels 12;	Gaps 5;

[illegible]

Db 185 NALPGLTATLSN----VDTFKEESRQKÖLKKIPKAFGRPEEYAAAAAALVSEASYT 240

QY 265 GETVVGGGT---PSRL 278
| | : | | | | |
Db 241 GATLEFVDGCMTLPSQL 257

Search completed: May 8, 2002, 08:27:33
Job time: 192 sec


```

Db 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278
|||||
RESULT 2
Q9NV08 PRELIMINARY: PRT: 278 AA.
AC Q9NV08:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE CDNA FLJ11008 FIS, CLONE PLACE1003100, MODERATELY SIMILAR TO HEP27
PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RT "NDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AK001870; BAA91953.1;
DR InterPro: IPR002198; Adh_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_C2.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 278 AA; 29509 MW; 6B0096BBA89152A0 CRC64;

Query Match 99.7%; Score 1395; DB 4; Length 278;
Best Local Similarity 99.6%; Pred. No. 3.6e-96;
Matches 277; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHKAGLGLCARANNVMASSGTRDPLANKVALYASTDGGFAIARRLAODGAHV 60
|||||
Db 1 MHKAGLGLCARANNVMASSGTRDPLANKVALYASTDGGFAIARRLAODGAHV 60
QY 61 VSSRKQONVDQAVATLOGEGLSVTGTVCHVKAEDRELVATVAVKLGIDILVNAVN 120
|||||
Db 61 VSSRKQONVDQAVATLOGEGLSVTGTVCHVKAEDRELVATVAVKLGIDILVNAVN 120
QY 121 PFGSINDVTEEVWVKTLIDINAKAPALMTKAVVPEMEKRGGSVIVSSIAFSPGFS 180
|||||
Db 121 PFGSINDVTEEVWVKTLIDINAKAPALMTKAVVPEMEKRGGSVIVSSIAFSPGFS 180
QY 181 PYNSKALGLITLTLEIAPRIRVNCIAPGLIKTSFSRMLMDKEESKMETLRIR 240
|||||
Db 181 PYNSKALGLITLTLEIAPRIRVNCIAPGLIKTSFSRMLMDKEESKMETLRIR 240
QY 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278
|||||
Db 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278

RESULT 3
Q9H3N5 PRELIMINARY: PRT: 260 AA.
AC Q9H3N5:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.

```

```

GN HUMNRD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Furukawa A., Onishi T., Huang D., Araki N., Ichikawa Y.;
RT "cDNA cloning and characterization of peroxisomal short-chain
RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AB045131; BAB18775.1;
DR InterPro: IPR002198; Adh_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_C2.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 260 AA; 27572 MW; 0BD2E0C0D2E37D08 CRC64;

```

```

Query Match 93.0%; Score 1301; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 3.2e-89;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MASSGTRDPLANKVALYASTDGGFAIARRLAODGAHVYSSRKQONVDQAVATLOG 78
|||||
Db 1 MASSGTRDPLANKVALYASTDGGFAIARRLAODGAHVYSSRKQONVDQAVATLOG 78
QY 79 EGLSTGTGTVCHVKAEDRELVATVAVKLGIDILVNAVNPFSGSINDVTEEVWVKTL 138
|||||
Db 61 EGLSTGTGTVCHVKAEDRELVATVAVKLGIDILVNAVNPFSGSINDVTEEVWVKTL 120
QY 139 DINAKAPALMTKAVVPEMEKRGGSVIVSSIAFSPGFSPIVNSKALGLITLAI 198
|||||
Db 121 DINAKAPALMTKAVVPEMEKRGGSVIVSSIAFSPGFSPIVNSKALGLITLAI 180
QY 199 ELAPRINIVNCIAPGLIKTSFSRMLMDKEESKMETLRIRRIGEPEDCAGIVSFLCSE 258
|||||
Db 181 ELAPRINIVNCIAPGLIKTSFSRMLMDKEESKMETLRIRRIGEPEDCAGIVSFLCSE 240
QY 259 DASVITGETVYVGGTSPSRL 278
|||||
Db 241 DASVITGETVYVGGTSPSRL 260

RESULT 4
Q95162 PRELIMINARY: PRT: 260 AA.
AC Q95162:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE PEROXISOMAL SHORT-CHAIN ALCOHOL DEHYDROGENASE.
GN SCAD-SRL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fransen M., Van Veldhoven P.J., Subramani S.;
RT "Identification of two novel mammalian proteins with a C-terminal
RT peroxisomal targeting signal by using the pvi-phase display
RT technology.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AF044127; AAD02292.1;

```

DR HSP: P50163; 2AE1.
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR Oxidoreductase.
 SQ SEQUENCE 260 AA; 27602 MW; 83687CACD82B9BC5 CRC64;

Query Match 92.4%; Score 1293; DB 4; Length 260;
 Best Local Similarity 99.2%; Pred. No. 1.3e-88;
 Matches 258; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 19 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQG 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258
 DB 181 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 240
 QY 259 DASITGETVVGCGTPSRL 278
 DB 241 DASITGETVVGCGTPSRL 260

RESULT 5
 Q9GKX2 PRELIMINARY; PRT; 260 AA.
 AC Q9GKX2:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.
 GN RABROR.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER.
 RA Fukukawa A., Ohnishi T., Huang D., Araki N., Ichikawa Y.;
 RT "cDNA cloning and characterization of peroxisomal short-chain
 RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
 CC FAMILY.
 DR EMBL: AB045133; BAB18777.1; .
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR Oxidoreductase.
 KW SEQUENCE 260 AA; 27429 MW; 5B0585B58911B90C CRC64;

Query Match 80.3%; Score 1123; DB 6; Length 260;
 Best Local Similarity 85.4%; Pred. No. 5.6e-76;
 Matches 222; Conservative 18; Mismatches 20; Indels 0; Gaps 0;

QY 19 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQA 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258
 DB 181 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 240
 QY 259 DASITGETVVGCGTPSRL 278
 DB 241 DASITGETVVGCGTPSRL 260

RESULT 6
 Q9E0U4 PRELIMINARY; PRT; 260 AA.
 AC Q9E0U4:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.
 GN MOURDR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57 BLACK; TISSUE-LIVER.
 RA Fukukawa A., Ohnishi T., Huang D., Araki N., Ichikawa Y.;
 RT "cDNA cloning and characterization of peroxisomal short-chain
 RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
 CC FAMILY.
 DR EMBL: AB045132; BAB18776.1; .
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR Oxidoreductase.
 KW SEQUENCE 260 AA; 27726 MW; 005C5F19AC539CF6 CRC64;

Query Match 77.6%; Score 1086; DB 11; Length 260;
 Best Local Similarity 81.5%; Pred. No. 3.1e-73;
 Matches 212; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 19 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPLANKVALYASTDGIQFAIARLAODGAHVYSSRKQONVDAVATLQG 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAFSPSPGFSPPYNSKTAALLGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258

-1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC EMBL: AK017274; BAB30665.1; -
 DR MGD: MGI:1918662; 5430405KRAK.R.
 DR InterPro: IPR002198; Adh_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR OXIDOREDUCTASE.
 KW OXIDOREDUCTASE.
 SO SEQUENCE 282 AA; 30070 MW; BB279D84DE7A4DF3 CRC64;

Query Match 56.3%; Score 787; DB 11; Length 282;
 Best Local Similarity 60.7%; Pred. No. 6.2e-51;
 Matches 159; Conservative 32; Mismatches 61; Indels 10; Gaps 2;

QY 16 SVRMASSGNTRRDPLANKVALVASTDGGIGFAIRRLADGAAHVYSSRKQONVDQAVAT 75
 DB 21 SARRRSKTADENRSLAGVAIVTGSTRGFAIRRLADGAAHVYSSRKQONVDQAVAT 80
 QY 76 LOGEGLSTGVCHVCAKEDRERLVAATKLGIDILVNAVPEFGSINDVTEEMD 135
 DB 81 LKEEGLSTGVCHVCAKEDRERLVAATKLGIDILVNAVPEFGSINDVTEEMD 140
 QY 136 KTLIDNVAPALMTKAVVPEMERKGGSSVYVSSIAAASPFGSPYNSKATLLGLTPT 195
 DB 141 KILDNVAVSPALLSKVLPYEMENRRGGSVIVSSGAAVVPKLVGYNTSKTALLGLCKS 200
 QY 196 LAIEAPRIRIRVNCIAPGLIKTSFSRMLMOKERE-----ESMKETLRIRRCPEPDCAG 250
 DB 201 LAVELAPRGIRVNCIAPGLIKTDFL-----REKTPMNMALPDMNKIFGVKRLGEPEECAG 255
 QY 251 IVSFLCSDDASTYTGTEVVG 272
 DB 256 IVSFLCSDDASTYTGTEVVG 277

RESULT 10
 Q9D2U3 PRELIMINARY; PRT; 216 AA.
 AC Q9D2U3;
 DT 01-JUN-2001 (TREMUREL. 17, Created)
 DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
 DE 01-JUN-2001 (TREMUREL. 17, Last annotation update)
 GN DNA SEGMENT, CHR 14, UNIVERSITY OF CALIFORNIA AT LOS ANGELES 2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=CEREBELLUM;
 RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukuishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Katsukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 RA Knehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,

RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC EMBL: AK018788; BAB31411.1;
 DR MGD: MGI:90169; D14UC12.
 DR InterPro: IPR002198; Adh_short.
 DR Pfam: PF00106; adh_short; 1.
 DR OXIDOREDUCTASE.
 KW OXIDOREDUCTASE.
 SO SEQUENCE 216 AA; 23178 MW; 88FC5B46D020D3E0 CRC64;

Query Match 51.2%; Score 716; DB 11; Length 216;
 Best Local Similarity 76.2%; Pred. No. 8.1e-46;
 Matches 147; Conservative 15; Mismatches 19; Indels 12; Gaps 2;

QY 1 MHKAG-LLGICAAVSVRASSGNTRRDPLANKVALVASTDGGIGFAIRRLADGAAHV 59
 DB 1 MOKAGRLGGMTQAMMSVMASSGLTRRNPLSNKVALVASTDGGIGFAIRRLADGAAHV 60
 QY 60 VVSRQONVDQAVATLOGEGLSTGVCHVCAKEDRERLVAATKLGIDILVNAV 119
 DB 61 VVSRQONVDQAVATLOGEGLSTGVCHVCAKEDRERLVAATKLGIDILVNAV 120
 QY 120 NPEFGSINDVTEEMKKTIDINKAPALMTKAVVPEMERKGGSSVYVSSIAA 174
 DB 121 NPEFGSINDVTEEMKKTIDINKAPALMTKAVVPEMERKGGSSVYVSSIAA 180
 QY 175 -----PSPGFSF 181
 DB 181 TCRSFHPTSTPP 193

RESULT 11
 Q9VRJ4 PRELIMINARY; PRT; 317 AA.
 AC Q9VRJ4;
 DT 01-MAY-2000 (TREMUREL. 13, Created)
 DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
 DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
 DE CG10672 PROTEIN.
 GN CG10672.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle E.K., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostalin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RESULT	12	
016619		
ID	016619	PRELIMINARY; PRT; 257 AA.
AC	016619;	
DT	01-JAN-1998 (TREMBLrel. 05	Created)
DT	01-JAN-1998 (TREMBLrel. 05,	last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17,	last annotation update)
DE	F36H9.3 PROTEIN.	
GN	F36H9.3.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea;	Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloiderinae;	Caenorhabditis.
OX	NCBI_Taxid=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
CC	STRAIN=BRISTOL N2;	

RESULT	13
Q93790	
ID	Q93790
AC	Q93790;
DT	01-FEB-1997 (TReMBLrel. 02.; created)
DT	01-MAY-1999 (TReMBLrel. 10.; last sequence update)
DT	01-JUN-2001 (TReMBLrel. 17.; last annotation update)
DE	F54PF3_4 PROTEIN.
CN	F54PF3_4
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea; Rhabdilitidae; Pelodierinae; Caenorhabditis.


```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
CC FAMILY.
DR EMBL: 279696; CAB01974.1; -.
DR EMBL: 281592; CAB01974.1; JOINED.
DR EMBL: 281592; CAB04734.1; -.
DR EMBL: 279696; CAB04734.1; JOINED.
DR HSSP: P50163; ZAEI.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00678; adh_short_C2.1.
DR Pfam: PF00678; adh_short_C2.1.
DR PRINTS: PR00080; SDRFAMILY.
DR Oxidoreductase.
KM SEQUENCE 260 AA: 27590 MW: E43FD36F5EDBA7F4 CRC64;

Query Match
Best Local Similarity 43.9%; Score 614; DB 5; Length 260;
Matches 120; Conservative 56; Mismatches 70; Indels 4; Gaps 2;

OY 33 KVALVTASTDGIQFAIRLAODGAHVYSSRKQONVDQAVATLOGEGLS-VTGTVCYHG 91
DB 11 KVAITTAARKGIGLAIAEFLDEGASVYIGSRQKNDVAIEYLNKGLTKVAAGIAGHTA 70
OY 92 KAEDRERLVATVATKLHGIDILVNAANVPFGSIMDVTAEVMDKTLIDINKAPALMTRA 151
DB 71 STDDQKILVDFLQKFGKINILVNNHGINPARGHILEVSDQYWDKLFEEVNVAAFGMTL 130
OY 132 VYPEMEKRGSGSVYVSSIAFSPSPGSPYVSKTALLGLTKTALIELAPNRIRVNCIA 211
DB 131 VAPHTAKBEGGAIIFNASSAYKSPPGIAAYGVTKTTGLTRALAMGLAKONIRVNGIA 190
OY 212 POLIKTSFRMLAM--MDKEEESMKETLIRRLGPEPDCAIVSFLCSEDAASYITGETV 268
DB 131 PEVITKMSQVLDGDEGDAEKELTIOETALORLGVPPDCACTVAYLASDSSYITGETV 250
OY 269 VVGCGTPSRL 278
DB 251 ITAGGVQARL 260

RESULT 14
OY 09S9W2 PRELIMINARY; PRT; 254 AA.
AC 09S9W2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE T1J24.9 PROTEIN (AT4G0530 PROTEIN).
GN T1J24.9 OR AT4G0530.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA All J., Bauer C., Nguyen C., Duckels G.;
RL "The sequence of A. thaliana T1J24."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Washu;
RL "The A. thaliana Genome Sequencing Project."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
CC FAMILY.
DR EMBL: AF147263; AAD48959.1; -.
DR EMBL: AL161503; CAB81095.1; -.
DR HSSP: P29132; IDFI.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00678; adh_short_C2.1.
DR Pfam: PF00678; adh_short_C2.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
DR Oxidoreductase.
KM SEQUENCE 254 AA: 26765 MW: 54297E4D1D095372 CRC64;

Query Match
Best Local Similarity 42.1%; Score 588.5; DB 10; Length 254;
Matches 126; Conservative 50.6%; Pred. No. 3.1e-36;
Mismatches 85; Indels 3; Gaps 2;

OY 30 LANKVALVASTDGIQFAIRLAODGAHVYSSRKQONVDQAVATLOGEGLSVTGTCH 89
DB 9 LEQKVAIVTASTGIGFIFGTERGLEGASVYSSRKQONVDVAIKSGIDAYGVCH 68
OY 90 VGKAEDRERLVATVATKLHGIDILVNAANVPFGSIMDVTAEVMDKTLIDINKAPALMT 149
DB 69 VSNQQRRLNVEKTVKYGKIDILVCAANSTDPILSSKEVADKLMEINVKSSILL 128
OY 150 KAVYPEMEKRGSGSVYVSSIAFSPSPGSPYVSKTALLGLTKTALIELAPNRIRVNC 209
DB 129 QDMAPHEK--GSSVIFITSIGFSPQAMAMGVTKTALLGLTKLAEMAP-DTRVNA 185
OY 210 LAPGLIKTSFRMLAMMDKEESAKETLIRRLGPEPDCAIVSFLCSEDAASYITGETV 269
DB 186 VAPGFVPIHFASFTIGTSSEVREGIEKTLNRLIGTGMMAAAAFASDSSYITGETV 245
OY 270 VVGCGTPSRL 278
DB 246 VAGGMP SRL 254

RESULT 15
OY 09DCL1 PRELIMINARY; PRT; 141 AA.
AC 09DCL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE DNA SEGMENT, CHR 14, UNIVERSITY OF CALIFORNIA AT LOS ANGELES 2.
GN D14UCLA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085650; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,

```

A

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:15:25 ; Search time 12.54 Seconds

(without alignments)
498.877 Million cell updates/sec

Title: US-09-866-034-2

Perfect score: 1399
Sequence: 1 MHKAGLLGLCARANNSVMA.....DASTITGETVYVGGTPTSSL 278

Scoring table: BIOSUM62
Gap 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 21252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	24.9	246	3	US-09-238-481-2
2	335.5	24.0	313	4	US-09-413-814-9
3	328.5	23.5	315	3	US-08-793-035-9
4	328.5	23.5	315	3	US-08-793-035-10
5	324.5	23.2	333	1	US-08-440-856A-4
6	324	23.2	262	4	US-09-363-189B-6
7	321	22.9	247	1	US-08-241-766-13
8	313	22.4	244	1	US-08-375-962B-13
9	313	22.4	244	2	US-08-562-114B-13
10	313	22.4	244	4	US-08-729-594A-13
11	312.5	22.3	256	1	US-08-594-808B-7
12	306.5	21.9	337	3	US-08-440-856A-3
13	299.5	21.4	303	3	US-09-002-298-1
14	296.5	21.2	257	4	US-09-287-097-2
15	295	21.1	244	1	US-08-762-129-1
16	295	21.1	244	1	US-09-090-567-2
17	295	21.1	271	2	US-07-637-865-2
18	294	21.0	244	1	US-08-762-129-3
19	279	19.9	246	6	5229279-7
20	272.5	19.5	255	4	US-08-815-225-4
21	270	19.3	251	3	US-08-822-322-9
22	270	19.3	251	4	US-09-466-109-9
23	270	19.3	252	3	US-08-822-322-8
24	270	19.3	252	4	US-09-466-109-8
25	266	19.0	244	1	US-08-762-129-4
26	258.5	18.5	283	4	US-09-367-012-1
27	253	18.1	273	6	5512669-4

28	248	17.7	359	1	US-08-440-856A-8	Sequence 8, Appl
29	246.5	17.6	248	4	US-09-385-028-11	Sequence 11, Appl
30	241	17.2	335	3	US-09-002-298-7	Sequence 7, Appl
31	237	16.9	295	3	US-09-002-298-5	Sequence 5, Appl
32	230.5	16.5	222	3	US-09-109-705-2	Sequence 2, Appl
33	226	16.2	335	3	US-09-002-298-6	Sequence 6, Appl
34	224	16.0	263	6	5229279-4	Patent No. 5229279
35	221	15.8	335	3	US-09-109-205-19	Sequence 19, Appl
36	217.5	15.5	231	2	US-08-336-198C-7	Sequence 7, Appl
37	217.5	15.5	285	4	US-09-026-482B-2	Sequence 2, Appl
38	197.5	14.1	327	1	US-08-375-962B-12	Sequence 12, Appl
39	197.5	14.1	327	2	US-08-562-114B-12	Sequence 12, Appl
40	197.5	14.1	327	4	US-08-729-594A-12	Sequence 12, Appl
41	191.5	13.7	261	4	US-08-815-225-2	Sequence 12, Appl
42	191.5	13.7	261	4	US-08-815-225-3	Sequence 2, Appl
43	183.5	13.1	388	4	US-08-980-832-41	Sequence 3, Appl
44	173	12.4	325	3	US-08-581-148C-4	Sequence 41, Appl
45	170.5	12.2	309	3	US-09-109-205-1	Sequence 4, Appl
						Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-238-481-2
Sequence 2, Application US/09238481
Patent No. 6110704
GENERAL INFORMATION:
APPLICANT: Huang, Jiansheng
APPLICANT: McDevitt, Damien
TITLE OF INVENTION: Fabg
FILE REFERENCE: GMI0192
CURRENT APPLICATION NUMBER: US/09/238,481
CURRENT FILING DATE: 1999-01-28
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 246
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-238-481-2

Query Match 24.9%; Score 348; DB 3; Length 246;
Best Local Similarity 32.6%; Pred. No. 3.8e-30;

Matches 79; Conservative 57; Mismatches 102; Indels 4; Gaps 3;

QY	33	KVALVTASTDGTGFAIARLAODGAHVVS-SRKOONVQDAVATLOGELSTGTGCHVG	91
DB	5	KSALVTGASRGISIAQLAEEGVNAVNVAGSKEKAAVEETKAKGVDSFAIOANVA	64
QY	92	KAEDEBRLVATVAKLHGIDIIIVSNAANPFEGSIMDYTEEYVDRITLDINAKAPALMTKA	151
DB	65	DADDEKAMKEVVSQFSGSDVAVNNAGITR-DNLMRMKEQMDVIDITNLKGVENCIOK	123
QY	152	VPEMEKRGGSVTVIAAFSPGSPFVNVSTALIGLTKTAIEIAPNINVCIA	211
DB	124	ATPQLKRSQAIINLSSVGAIVNGQANVATAGVIGLTKSAAREIASRGITVNAVVA	183
QY	212	PGLIKTSFRLMMDKEESKFTLRIRLGEPEDCAGIVSFLSEDSASTTGTYYVG	271
DB	184	PGFIYSMDTDL-SDELKEQMLTOIPLARFGODDIANTVAFILASDAKAKYTGOTIHVN	241
QY	272	GG 273	
DB	242	GG 243	
RESULT 2			
	US-09-413-814-9		
	Sequence 9, Application US/09413814		
	Patent No. 6225064		
	GENERAL INFORMATION:		

```

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoeller, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO: 9
LENGTH: 313
TYPE: PRF
ORGANISM: Sorangium cellulosum
US-09-413-814-9

```

```

Query Match      24.0%: Score 335.5; DB 4; Length 313;
Best Local Similarity 33.6%: Pred. No. 1.3e-28;
Matches 83; Conservative 50; Mismatches 105; Indels 9; Gaps 4;

```

```

QY 30 LANKVALVASTDGGIGFAIARLADGAAHVVSRRKQON-VDAVAVTLOGEGLSVGTVC 88
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 50 LAGRVALLVSSRGIGKALALRLAEGADYAVVHNKDAEPTAELIRALGRTRVQA 109
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 89 HVGKADRERLVATANKLHGIDILVNAVNEFFESIMDVTEEWKTLIDINVKAPALM 148
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 110 DTRPNAALFSSVPAQGLPIDILVNVG-DEFFKPLAAMDDEMRNVMSLSVHL 168
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 TRAVVPEMERGGGSVYI--SSIAFSPSPGSPYVNSKTKLTLGLTKLATELARINR 206
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 169 CAAVARMORSGRIINIGLSPYTAIRGAPVNAATSTAKTGLVLTLSLATEEAPHGIL 228
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 207 VNCGLAPGLITFSRMLMDKEKESMKETLIRLIGPEPDAGIVSFCSEDAVITGE 266
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 229 VNCVSPGLIDNGY-----LPPAQKEMERRVPMGRIGRASEVADANAFLASDRASVSGA 283
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 267 TVVVGCG 273
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 284 NIAVAGG 290

```

```

RESULT 3
US-08-793-035-9
Sequence 9, Application US/08793035
Patent No. 6011201
GENERAL INFORMATION:
APPLICANT: Slabas, Antoni R.
APPLICANT: White, Andrew
APPLICANT: Chase, Diane
APPLICANT: Elborough, Keiran
APPLICANT: Fentem, Phillip A.
TITLE OF INVENTION: B-ketocyl ACP Reductase Genes From
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,035
FILING DATE: 28-JUL-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9414622.2
FILING DATE: 20-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB95/01678
FILING DATE: 17-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: MOBR:132
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713.787.1400
TELEFAX: 713.787.1440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-793-035-9

```

```

Query Match      23.5%: Score 328.5; DB 3; Length 315;
Best Local Similarity 33.1%: Pred. No. 7.9e-28;
Matches 81; Conservative 53; Mismatches 100; Indels 11; Gaps 5;

```

```

QY 34 VALVASTDGGIGFAIARLADGAAHVYV---SSRKOQNVDAVATLOGEGLSVGTVC 89
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 74 VVVVTSRIGSRIKALSLGKACQVYVNAKAEKEVSKOIEYGGATFFG---D 130
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 90 VGEADRERLVATANKLHGIDILVNAVNEFFESIMDVTEEWKTLIDINVKAPALMT 149
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 VSKADVEMMKTAIDAMGTIDVYVNNAGITR-DTLIRKKSQOMDEVIDNLTVGLCT 169
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 150 KAVVPEMERGGGSVYIVSSIAFSPSPGSPYVNSKTKLTLGLTKLATELARINRVC 209
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 190 QAAFTKIMMKRKCRITINISVGLIGNIGQANAAKAGVIGFSKTAAREGASRNINNV 249
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 210 LAPGLITFSRMLMDKEKESMKETLIRLIGPEPDAGIVSFC-SEDAVITGETV 268
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 250 VCPGFASDMTKLGEDMEKK--ILGTIPGRIGPEDEVAGLVERFLASPAASYITGQAF 307
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 269 VVVGCG 273
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 308 TIDGG 312

```

```

RESULT 4
US-08-793-035-10
Sequence 10, Application US/08793035
Patent No. 6011201
GENERAL INFORMATION:
APPLICANT: Slabas, Antoni R.
APPLICANT: White, Andrew
APPLICANT: Chase, Diane
APPLICANT: Elborough, Keiran
APPLICANT: Fentem, Phillip A.
TITLE OF INVENTION: B-ketocyl ACP Reductase Genes From
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX

```

```

: COUNTRY: US
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/793,035
: FILING DATE: 28-JUL-1997
: CLASSIFICATION: 800
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: GB 9414622.2
: FILING DATE: 20-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB95/01678
: FILING DATE: 17-JUL-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Kammerer, Patricia A.
: REGISTRATION NUMBER: 29,775
: REFERENCE/DOCKET NUMBER: MOBT:132
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713.787.1400
: TELEFAX: 713.787.1440
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 315 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-793-035-10

```

```

Query Match      23.5% Score 328.5; DB 3; Length 315;
Best Local Similarity 33.1%; Pred. No. 7.9e-28;
Matches 81; Conservative 53; Mismatches 100; Indels 11; Gaps 5;

: 34 VALVASTGIGFATARRLAODGAHVY-----SSRQONVDAVATLQEGSLVTGTGCH 89
: 74 VVVVVGASGICATATLSIGKAGCKVLYVYASAKAEVSKQIEAYVGGQATFFGG--D 130
: 90 VKAEDRELVATAVAKLHGIDILVSNAAVNPFEGSIMDVEEVMKTDIDINVKAPALMT 149
: 131 VKKADVEAMKTAIDAMCTIDVYVNNAGITR-DTLLIMKKSQNDVLDNLTGFLCT 189
: 150 KAVVEEMERGGGVYVYSIAFSPSPFYNSKTAALLGLTATLATELAPRNIRIVNC 209
: 190 QAAATIMMKRRGRINIINIASVGLNIGIQANYAAKAGVIGFSKTAAREGASRNINAVY 249
: 210 LAPGLIKTSFSEMLMDKEKESMKETLIRLGEPEDCAGIVSFLC-SEDSYITGETV 268
: 250 VCPGSIASMTAKLGEDMEKK--ILGTIPLGRYGPEDVAGLVEFLALSPASYSITGOAF 307
: 269 VVGGG 273
: 308 TIDCG 312

```

```

: RESULT 5
: US-08-440-856A-4
: Sequence 4, Application US/08440856A
: Patent No. 5750873
: GENERAL INFORMATION:
: APPLICANT: DELLAPORTA, STEPHEN L.
: TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 PENNSYLVANIA AVE. N.W.
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA

```

```

: ZIP: 20037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440,856A
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: MILLMAN, ROBERT A.
: REGISTRATION NUMBER: 36,217
: REFERENCE/DOCKET NUMBER: 05463-20001.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1517
: TELEFAX: (202) 887-0763
: TELETYPE: 706141
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 333 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-440-856A-4

```

```

Query Match      23.2% Score 324.5; DB 1; Length 333;
Best Local Similarity 32.9%; Pred. No. 2.4e-27;
Matches 102; Conservative 37; Mismatches 108; Indels 63; Gaps 11;

: 13 ANNSVRMASSGMRDPLANKVALYASTDIGFATARRLAODGAHVYSSRQONVDA 72
: 37 AMDSPNGCAPMPMKR--LECKVAIVTGARGIGEAIVRLVYKGANVYA-----DIDDA 89
: 73 VALQEGSLV-----TGIV-CHVKAEDRELVATAVAKLHGIDILVSNAAVNPFEG-- 124
: 90 A--GEALAAALGPVHGFRCVSVSEEDVERAVENAVARYGRDLVNNAGV---LGRQ 142
: 125 -----SIMDVEEVMKTDIDINVKAPALMTKAVPEMEKRGSGSVYVYSIAFSPSPG 179
: 143 TRAKKILSTPDAGEFPRVRLVNALGALGMKHAALAMTORRAGSITSVASVAGVLGLGP 202
: 180 SPYNSKTAALLGLTATLATELAPRNIRIVNCLAPGLIKTSFSEMLMM----- 225
: 203 HATYASKHAIIVGLTKNACELAGHIGIRVNCISPFVATPLIANNQGHDASTADADAD 262
: 226 -----DKEKESKKEFLR-----IRLGEPEDCAGIVSFLCSEDSYITGETV 268
: 263 IDLDIAPSDQVEKMEVEVYRGATLKGATLR---PRDIAEALPLIASDSRYSIGHNL 318
: 269 VVGGG-TPSR 277
: 319 VVDGCVTTSR 328

```

```

: RESULT 6
: US-09-363-189B-6
: Sequence 6, Application US/09363189B
: Patent No. 6242228
: GENERAL INFORMATION:
: APPLICANT: SUGIYAMA, MASAKAZU
: APPLICANT: TONOUCHI, NAORO
: APPLICANT: SUZUKI, SHUNICHI
: APPLICANT: YOKOZAKI, KENZO
: TITLE OF INVENTION: XYLTOL DEHYDROGENASE OF ACETIC ACID BACTERIA AND GENE THERE
: FILE REFERENCE: 0010-1024-0
: CURRENT APPLICATION NUMBER: US/09/363,189B
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: JP10-216047
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patent In Release #1.0

```

```

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```


TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 244 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 NAME/KEY: E.coli 3-oxoacyl[acyl carrier protein]reductase
 NAME/KEY:
 US-08-729-594A-13

Query Match 22.4%; Score 313; DB 4; Length 244;
 Best Local Similarity 32.8%; Pred. No. 2,6e-26;
 Matches 81; Conservative 53; Mismatches 95; Indels 18; Gaps 6;

33 KVALVASTDGIQAIIRRLAODGAHVVSRRKQNDQAVT--LQEGSLVGTGVCH 89
 6 KIALVTSASRGIAETLAARGK-VIGATSENQAQAIISDYLKANGKGLMLNTV--- 61
 90 VGRAEDRERLVATAVKLH---GGIDILVNAAVNPFSGSIMDVTEEVWDTLIDINAKA 146
 62 -----DPASIESYLEKIRAEFGEDILVNNAGITR-DNLMRKDEKNDIETNLSVF 115
 147 LMTKAVPEMEKRGSGSVIVSSIAFSPSPSPYVSKTALLGLTKTLAIELAPNIR 206
 116 RLSKAVARMMKKRHRIITIGSVGTMGNGQANAAKAGLIGFSKLAREVASGIT 175
 207 VNCLAPGLITFSRMLMMDKEESKKEITLIRIRLGEPEDCAGIVSFLCEDASYTGE 266
 176 VNVVAPGFIETDMTRAL--SDQAGILAOVPAGRLGAGOEIANAVALASDEAYITGE 233
 267 TVVVGCG 273
 234 TLHVNGG 240

RESULT 11

US-08-594-808B-7
 Sequence 7, Application US/08594808B
 Patent No. 5804423

GENERAL INFORMATION:
 APPLICANT: Klasen, Ralf
 APPLICANT: Bringer-Meyer, Stephanie
 APPLICANT: Salm, Hermann
 APPLICANT: Hollenberg, Cornelies P
 TITLE OF INVENTION: MICROBIOLOGICAL METHOD OF MAKING
 TITLE OF INVENTION: 5-KETOGLUCONATE
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: The Firm of Karl F. Ross, PC
 STREET: 5676 Riverdale Ave.
 CITY: Bronx
 STATE: New York
 COUNTRY: USA
 ZIP: 10471-0900

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/594,808B
 FILING DATE: 07-FEB-1995
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Myers, Jonathan E
 REGISTRATION NUMBER: 26,963
 REFERENCE/DOCKET NUMBER: 19893
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (718) 884-6600
 TELEFAX: 718/601-1099

TELEX: 620428
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 256 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-594-808B-7

Query Match 22.3%; Score 312.5; DB 1; Length 256;
 Best Local Similarity 31.2%; Pred. No. 3,2e-26;
 Matches 77; Conservative 53; Mismatches 110; Indels 7; Gaps 3;

30 LANKVALVASTDGIQAIIRRLAODGAHVVSRRKQNDQAVTLQEGSLVGTGVCH 89
 9 LSGARLVATGASRGIGTLKGLARYAEVVLNGRNSESIDSAGSGFEAREGLKASTAVFD 68
 90 VGRAEDRERL--VATAVKLHGGIDILVNAAVNPFSGSIMDVTEEVWDTLIDINAKA 146
 69 V---TDQDAVIDGYAAIERDMGPIDILINNAGIOR-RAPLEEFSRKDDMLSTNNAVF 124
 147 LMTKAVPEMEKRGSGSVIVSSIAFSPSPSPYVSKTALLGLTKTLAIELAPNIR 206
 125 FVQGVAVARHMPRGKIVNICSVOSELARGIAPTATGAVKNLTKMATDMGRGLQ 164
 207 VNCLAPGLITFSRMLMMDKEESKKEITLIRIRLGEPEDCAGIVSFLCEDASYTGE 266
 185 INGLAPGYFTEMTERLVADEFTDMLCKRTPAQRWGOVELVGAVALSSRRASSFVNGQ 244
 267 TVVVGCG 273
 245 VLMWDGG 251

RESULT 12

US-08-440-856A-3
 Sequence 3, Application US/08440856A
 Patent No. 5750873

GENERAL INFORMATION:
 APPLICANT: DELLAPORTA, STEPHEN L.
 TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
 TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 2000 PENNSYLVANIA AVE. N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,856A
 FILING DATE: 15-MAY-1995
 CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
 NAME: MILLMAN, ROBERT A.
 REGISTRATION NUMBER: 36,217
 REFERENCE/DOCKET NUMBER: 05463-20001.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 887-1517
 TELEFAX: (202) 887-0763
 TELEX: 706141

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 337 amino acids
 TYPE: amino acid
 STRANDEDNESS: single

OY 148 MTKAVPEMEKRGSGSVIVSIAAFSPSPGFSPYVNSKTALLGLTKTALIELAPRNIRV 207
 Db 118 TTAALDGMKAGKGRIVNATASAHGLTASPKSAVYAKHGVTGTALLETAGKGTIC 177
 OY 208 NCLAPGLIKTSF-----SHLMWMDKEESMKETL-----RIRLGEPCAGIVSFLCS 257
 Db 178 NALCPGVLTPLVEAQPDDMKAHMDMRETVIRVMDROPSROFATTGIGTIVFLCS 237
 OY 258 EDASVITGETVYVGGG 273
 Db 238 GAADQITGTTISVDGG 253

RESULT 15

US-08-762-129-1
 ; Sequence 1, Application US/08762129
 ; Patent No. 5756299
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; TITLE OF INVENTION: A NOVEL HUMAN CARBONYL REDUCTASE
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/762,129
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0171 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 244 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Consensus
 ; CLONE: Consensus
 ; US-08-762-129-1

Query Match 21.1%; Score 295; DB 1; Length 244;
 Best Local Similarity 32.7%; Pred. No. 2.5e-24;
 Matches 81; Conservative 44; Mismatches 107; Indels 16; Gaps 5;

OY 30 LANKVALVASTDGTGFARLADGAAHVVSRRQONVDOAVATLOGEGLSVTGTCH 89
 Db 5 LAGRYLVYTGAGKIGRGTVQALHATGARVAVSRFQADLDLSVRCFG---IEPVCD 60
 OY 90 VKAEDRERLVTAVKLHGIGDILVSNAAV---NPFSGIMDVTEEVMDKTDIVAKAPA 146
 Db 61 LGDWATERALGVS---GPVLDLVNNAVALLOPF---LEVTKEAFDRSEFVNLRAVI 112

OY 147 LMTKAVPEMEKRG-GGSVIVSIAAFSPSPGFSPYVNSKTALLGLTKTALIELAPRNI 205
 Db 113 QVSQIVARGLIARGVPGAIIVVSSQCSORAVTNHNSVYICSTGALDMLTKVALELGPRTI 172
 OY 206 RVNCLAPGLIKTSFSSRMIMDKKEESMKETLIRIRLGEPCAGIVSFLCSEDASYITG 265
 Db 173 RVNAVNPFTVYVMTSMGQATWSDPHKAKTWNRIPIGKFAVEYHVYVNAIIFLSDRSQMTTG 232
 OY 266 ETVVVGGG 273
 Db 233 STLPEVGG 240

Search completed: May 8, 2002, 08:23:47
 Job time: 502 sec


```

1 TITLE OF INVENTION: chloritetracycline and tetracycline formation and cosmid
2
3 TITLE OF INVENTION: useful therein
4
5 NUMBER OF SEQUENCES: 1
6
7 CORRESPONDENCE ADDRESS:
8 ADDRESSER: American Cyanamid Company
9 STREET: One Cyanamid Plaza
10 CITY: Wayne
11 STATE: New Jersey
12
13 COUNTRY: USA
14
15 ZIP: 07470
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: floppy disk
19
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patent In Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/474,933
25
26 FILING DATE:
27
28 CLASSIFICATION: 435
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/125,468
32
33 FILING DATE: 22-SEP-1993
34
35 ATTORNEY/AGENT INFORMATION:
36
37 NAME: Tsevdos, Estelle J
38 REGISTRATION NUMBER: 31,145
39 REFERENCE/DOCKET NUMBER: 31,255-02
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: (201)831-3241
42 TELEFAX: (201)831-3305
43
44 INFORMATION FOR SEQ ID NO: 1:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 30001 base pairs
47 TYPE: nucleic acid
48 STRANDEDNESS: single
49 TOPOLOGY: linear
50
51 MOLECULE TYPE: DNA (genomic)
52
53 US-08-474-933-1

```

[illegible]

```

99  LeuValAlaThrAlaValLysLeuHisGlyGlyThrLeuSerLeuValSe 115
      |||||.....||| |||..... |||.....
8169 CTCTCGGGGGGGGGGACCGCTTCGGCGCGGCTGCGGTCA 8120
      |||||..... |||..... |||.....
115  rAsnAlaAlaValAsnProPhePheGlySerIleMetAspValThrGlu 132
      |||||..... |||..... |||.....
8119 CAAGCGCGGAGCGAAGCGG...GGTGGCGTCACTCCACCCCTCCAGCG 8073
      |||||..... |||..... |||.....
132 LuValTTPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMet 148
      |||||..... |||..... |||.....
8072 AGCTGTGGCAGAGCATGTGACACCAACTCACACGCTTCGCCGCTC 8023
      |||||..... |||..... |||.....
149 ThrLysAlaValAlaProGlu.....MetGlyLysArgGlyGlyGlySe 163
      |||||..... |||..... |||.....
8022 ACCCGGAGGCTCTCACACCGCGCGGCGATGAGCGCGCGGCAAGCGCG 7973
      |||||..... |||..... |||.....
163 rValValIleValSerSerIleAlaAlaPheSerProSerProGlyPheS 180
      |||||..... |||..... |||.....
7972 GATCATCAGGCTCCCTCCACCGCGCGGCAAGCGGTGTCCCGCTGGCG 7923
      |||||..... |||..... |||.....
180 erProTTPAsnValSerLysThrAlaLeuGlyLeuThrLysThrLeu 196
      |||||..... |||..... |||.....
7922 CCCCTACTCGGCTCCAGCGCGGCTCATCGCTTACCAAGGGCGCTG 7873
      |||||..... |||..... |||.....
197 AlaIleGluLeuAlaProAlaGluIleArgValAsnCysLeuAlaProG 213
      |||||..... |||..... |||.....
7872 GCCAAGGAACTGCCCCACACCGGCGACCGTCAACGCCGTGCCCGG 7823
      |||||..... |||..... |||.....
213 YLeuIleLysThrSerPheSer.....ArgMetL 223
      |||||..... |||..... |||.....
7822 CTACGTCGACACCGCGATGGCGCTGCCGCGTCCCGCAAGCGCGCGCA 7773
      |||||..... |||..... |||.....
223 euTPMetAspLysGluLysGlu.....GluSerMetLysGluThrLeu 237
      |||||..... |||..... |||.....
7772 CTGGGCGACACCGAGGAGGAGGTGCTCGCGCTTCGAGCAGAGATC 7723
      |||||..... |||..... |||.....
238 ArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIleValSerP 254
      |||||..... |||..... |||.....
7722 CGCGTCGGCGCTACTCCACCGCGGAGGAGGTGCGCGCATGTCGACT 7673
      |||||..... |||..... |||.....
254 eLeuCysSerGluAspAlaSerThrIleThrGlyGluThrValValG 271
      |||||..... |||..... |||.....
7672 CTGTACACACCGCGCGCGGTGCTCACCGCGCAGCGCATCAACGTCT 7623
      |||||..... |||..... |||.....
271 LysGlyGly 273
      |||||..... |||..... |||.....
7622 GCGGTGGC 7615

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-238-481-1

```

seq_documentation_block:
; Sequence 1, Application US/09238481
; Patent No. 6110704
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: Pabg
; FILE REFERENCE: GM10192
; CURRENT APPLICATION NUMBER: US/09/238,481
; CURRENT FILING DATE: 1999-01-28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-238-481-1

```

```

alignment_scores:
Quality: 348.00 Length: 242
Ratio: 1.989 Gaps: 3
Percent Similarity: 72.314 Percent Identity: 32.645

```

alignment_block:
US-09-866-034-2 x US-09-238-481-1
Align seg 1/1 to: US-09-238-481-1 from: 1 to: 741

```

33  LysValAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAl 49
      ||| |||||..... |||..... |||.....
13  AAGAGTGCTTAGTAACAGGTGATCAAGAGAAATGAGACTATATTCG 62
      |||||..... |||..... |||.....
49  aArgArgLeuAlaGluAspGlyAlaHisValValSer...SerArgL 65
      |||||..... |||..... |||.....
63  GTTCAATTTAGCAGAGAGAGATTAATGTAGCAGTAACATATGACAGCA 112
      |||||..... |||..... |||.....
65  YSGInGlnAsnValAspGluAlaValAlaThrLeuGlnGlyGlyLeu 81
      |||||..... |||..... |||.....
113  GCAAGAGAAAGCTGAAGCAAGTACTGCAAGAAATCAAAAGCTAAAGGT 162
      |||||..... |||..... |||.....
82  SerValThrGlyThrValLysHisValGlyLysAlaGluAspArgGlu 98
      |||||..... |||..... |||.....
163  GACAGTTTGGGATTCAGCAAAATGTCGCGATGCTGATGATTAAGC 212
      |||||..... |||..... |||.....
98  gLeuValAlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValS 115
      |||||..... |||..... |||.....
213  AATGATTAAAGAGTACTAGCCAAATTTGTTTGTATGATGTCTTACTAA 262
      |||||..... |||..... |||.....
115  eAsnAlaAlaValAsnProPhePheGlySerIleMetAspValThrGlu 131
      |||||..... |||..... |||.....
263  ATAATGCAAGTATTATCTCGC...CATTAATTATTAAATCGCTATGCAAG 309
      |||||..... |||..... |||.....
132  GluValTTPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMe 148
      |||||..... |||..... |||.....
310  CAAGAGGGGATGATGATTATGACACAACTTAAGAGGTATTTAACTG 359
      |||||..... |||..... |||.....
148  tThrLysAlaValAlaProGluMetGluLysArgGlyGlySerValY 165
      |||||..... |||..... |||.....
360  TATCCAAAAGACACACCAAAATGTTAAGCAACGTAAGTGCTATCA 409
      |||||..... |||..... |||.....
165  alIleValSerSerIleAlaAlaPheSerProSerProGlyPheSerPro 181
      |||||..... |||..... |||.....
410  TCAATTTATCAAGTGTGTTGGAGCCACTAGTAAATCGGAGCAAGCAAC 459
      |||||..... |||..... |||.....
182  TyrAsnValSerLysThrAlaLeuGlyLeuThrLysThrLeuAlaI 198
      |||||..... |||..... |||.....
460  TATGTTGCACAAAGAGAGGTGTTATGTTTAACTTAATCTGGCGCG 509
      |||||..... |||..... |||.....
198  eGluLeuAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeu 215
      |||||..... |||..... |||.....
510  TGAATTAGCATCTCGTGTATCACTTAATGCAAGTTCGACCTGGTTTA 559
      |||||..... |||..... |||.....
215  LysThrSerPheSerArgMetLeuThrMetLysLysGlyGlu 231
      |||||..... |||..... |||.....
560  TTGTTTCTGTATGACAGATCTTTA...AGTGTAGCTTAAAGAA 603
      |||||..... |||..... |||.....
232  SerMetLysGluThrLeuArgIleArgArgLeuGlyGluProLysPcy 248
      |||||..... |||..... |||.....
604  CAATGTTGACTCAAAATTCGTTAGCAGGTTTGGTCAACACAGATAT 653
      |||||..... |||..... |||.....
248  saIaGlyIleValSerPheLeuCysSerGluAspAlaSerThrIleThr 265
      |||||..... |||..... |||.....
654  TGGTAATACAGTAGCGTTTACATCAGCAAGCAAAATATATTTACAG 703
      |||||..... |||..... |||.....
265  LysLuhThrValValAlaGlyGly 273
      |||||..... |||..... |||.....
704  GTCAACCAATCCATGTAAATGTGGA 729

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-239-052-1

```

seq_documentation_block:
; Sequence 1, Application US/09239052
; Patent No. 6346395
; GENERAL INFORMATION:
; APPLICANT: Holmes, David J.

```

[illegible]

30/0962CTGCCGGGCCCATGTGCTGCTCACCGCACGCGAGGAGCTGCCGAC 3070913

71 GlnAlaValAlaThrLeuGlnGlyGlyGlyLeuSerValThrGlyThrVal 87

```

      3070912GAGGCTGCTCCACAGGTCCGGCCGACCGCCCTG.....GGCGTTGG 3070872
      87 LCYSHISVALGLYLSALAGLUAAPARGLUARGLEUVALATHRALAV 104
      3070871TGCACACCGCGTGCAGCAGATGCCCGCGCGTGTGTGACCTCCACC 3070822
      104 aLysLeuNHISGLYGLYLEASPILEUVALSERASNAIAALAVASN 120
      3070821TCGAGGCGCTTCGCGACGTGTGACATCTGTATCAACAGCGCGGACCAAC 3070772
      121 ProPhePheGlySerIleMetAspValThrGluGluValTrpAspIle 137
      3070771CCGCGTTACGCTCCGCTGCAGACAGACCGCCCGCTTCGCAAGAT 3070722
      137 rLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAlaValAlp 154
      3070721CTTCGACGTCAACCTGTGGCACCCTGTGATGTGACCTCCCTGCTGA 3070672
      154 roGluMetGluLysArgGlyGlyGlySerValIleValSerIle 170
      3070671CGGCGTGCATGGCGGACAGCGCGCGCGTGTCAACACCGCTCCATC 3070622
      171 AlaAlaPheSerProSerProGlyPheSerProTyrAsnValSerIle 187
      3070621GGCGGGATGCACGATGCCCGGCATGGCATGTACAAACCGCAACAGC 3070572
      187 rAlaLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaProArg 204
      3070571CGGCGTATCCAGTCCACAGCAAGCATGGCGCGTGTCAACCGGC 3070523
      204 snIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSer 220
      3070522..ATCCGGGTGATGGGATGTGTCCCGGGTGTCCACACGAGCTTACC 3070475
      221 ArgMetLeuTrpMetAspLysGlyLysGluLysSerMetLysGluThrIle 237
      3070474GAGCGCCCTGTGG.....AAGGACCAACAGAGATCCGTTGGCGCGACCAT 3070431
      237 uArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIleValSerP 254
      3070430TGGCGTTCGACGATCGGGAGCGCGCCGACATAGCAGCGCGGTCCGT 3070381
      254 heLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrValAlaVal 270
      3070380TCTGTGTTGGATGCGCGCAAGCTGATCCGCGCAACCATGATCATC 3070331
      271 GlyGlyGly 273
      3070330GACGCGCGGT 3070322
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
: Sequence 1, Application US/09103840A
: Patent No. 6294338
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103.840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv

```

```

US-09-103-840A-1
alignment_scores:
  Quality: 462.50      Length: 253
  Ratio: 2.473        Gaps: 4
  Percent Similarity: 73.913      Percent Identity: 39.130
alignment_block:
US-09-866-034-2 x US-09-103-840A-1/rev
Align seq 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529
21 SerGerglyMetThrArgAspProLeuAlaAsnLysValAlaLeuVal 37
30763766CAAGCCAAAGACTGCTACTGAT...CTGACCGCGCCGACCGCCATTAAT 3076330
37 rThAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArgLeuAlaG 54
3076329TACTGGCGCATCGCGAGCGATCGGGCTTCCATCGCCCAACAGCTGGCAG 3076280
54 lnaSprGlyAlaNHISValValValSerArgLysGlnGlnAsnValAsp 70
3076279CTCCGCGCGCCCATGTGTGCTCACCGCAGCAGAGAAAGTGGCCGAC 3076230
71 GluAlaValAlaThrLeuGlnGlyGlyLysSerValThrGlyThrVal 87
3076229GAGGCTGTCCACAGGTGCGCGACCGCGCCCTG.....GGCGTTGG 3076189
87 LCYSHISVALGLYLSALAGLUAAPARGLUARGLEUVALATHRALAV 104
3076188TGCACACCGCGTGCAGCAGATGCCCGCGCGTGTGTGTGACTCCACC 3076139
104 aLysLeuNHISGLYGLYLEASPILEUVALSERASNAIAALAVASN 120
3076138TCGAGCCCTTCGCGCAGCGTTGATCTCTGATCAACAGCGCGAACCAC 3076089
121 ProPhePheGlySerIleMetAspValThrGluGluValTrpAspIle 137
3076088CGGCTTACGCTCGCGTGTGAGCAGACACCGCCGCTTCCGCAAGAT 3076039
137 rLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAlaValAlp 154
3076038CTTGACGTCAACTGTGGCAGCGCTGATGTGACCTCGCTGCTGTA 3075989
154 roGluMetGluLysArgGlyGlySerValIleValSerIle 170
3075988CGCGTGCATGCGCGACGACGCGCGCGGTGTCAACACCGCTCCATC 3075939
171 AlaAlaPheSerProSerProGlyPheSerProTyrAsnValSerIle 187
3075938GCGGATGCACGATCCCGCGCATGGCATGTATCAACAGCGCAACAGC 3075889
187 rAlaLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaProArg 204
3075888CGCGCTGATCCACTCAACAGCAGCTGCGCGTGTGACCTTCCACCGCGC 3075840
204 snIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSer 220
3075839..ATCCGGGTGAATGCGATGTCTCCCGGGGTGTTCCACACGAGCTTGGC 3075792
221 ArgMetLeuTrpMetAspLysGlyGluLysSerMetLysGluThrIle 237
3075791GAGGCGCTGTGG.....AAGGACCAACAGAGATCCGTTGGCGCGCAT 3075748
237 uArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIleValSerP 254
3075747TGGCGTTCGACGATCGGGAGCGCGCGCATATGCAAGCGGTCCGT 3075698
254 heLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrValAlaVal 270
3075697TCTGTGTTGGATGCGCGCAAGCTGATCAACCGCGCAACCATGATCATC 3075648

```

271 Glyglygly 273

3075647GACGCGGT 3075639

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:

```

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "a" bases at various positions throughout the sequence
US-09-103-840A-2

```

alignment_scores:

```

Quality: 368.00      length: 284
Ratio: 1.957         Gaps: 8
Percent Similarity: 66.197      Percent Identity: 35.563

```

alignment_block:

US-09-866-034-2 x US-09-103-840A-2

Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765

```

10 CysAlaArgAlaTrp.....AsnSerValAr 18
|||||
2190895TGTCGCGCATTTGGCGGTTCGCGGATCCGAGACCGACGA 2190944
18 gMetAlaSerSerGlyMetThrArgArgSppProLeuAlaAsnLysVala 35
|||||
2190945CGTTGGCGCGCCGACGTTGAACCATCCAGAC...CTGGCCGCGCAAGGTGC 2190991
35 lAlaValAlaThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArg 51
|||||
2190992CGATCTGTCTGGCGCGCGCGCGCGGCGGATGCGTGGCGGTGGCGCGCA 2191041
52 LeuAlaGlnAspGlyAlaHisValValValSerSerArgLysGlnLnas 68
|||||
2191042CTTCGCGCGACGCGCTGCTGCTGCGCGACATCGATGCTGATGC 2191091
68 nValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValThrG 85
|||||
2191092CCGCGGAT...GCCGCGCGCCACCAAAATCGTCTGTGC...G 2191126
85 lYThrValCysHisValGlyLysAlaGlnAspArgLnu.....ArgLeu 99
|||||
2191127CAGCGCGCTCGCGGTTGACGTCAGCGACGACACATCATCGCCATG 2191176
100 ValAlaThrAlaValAlaLysLeuHisGlyGlyIleAspIleLeuValSerAs 116
|||||
2191177GTTCGACGCGCTGTGTCCGCGTTCGCGGCGGTGACACACTGTGTCGCCA 2191226
116 nAlaAlaValaAsnProphepGlySerIleMetAspValThrGlnGluV 133
|||||
2191227CCGCGGTGTC...GTTCATCTGTGCTTCGCTCATTCGACACACCGCTGCGAG 2191273

```

```

133 aITrPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThr 149
|||||
2191274ACTTCGATCGGCGTCATCCGATCATCTCCGCGCGCTCGCTGTCACCC 2191323

```

```

150 LysAlaValAlaProGluMetGlnLysArgGlyGlySerValValI 166
|||||
2191324AACGATCGCGGACCCCGGATATCTCAGCGCGCGGGGAGGACCATTTGCA 2191373

```

```

166 eValSerSerIleAlaAlaPheSerProSerProGlyPheSerProYra 183
|||||
2191374CCTGTCGCTTAGCGGCGCAGGTAGCGGTGGCGGACCGCCCATRAG 2191423

```

```

183 snValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 199
|||||
2191424GCGATGTCGACAGCGCGCATCATCCAGCTCAGCCGATCAGCCGCGCA 2191473

```

```

200 LeuAlaProAlaGlnAlaValAlaAspCysLeuAlaProGlyLeuIle 216
|||||
2191474CTGCGCTCGTGGCGCATCCCTCCACACGCTGCTGCGCCGATTCGCA 2191523

```

```

216 sThrSerPheSerArg.....MetLeuTrpMetAspLysGluLysGlu 231
|||||
2191524ACACCCGATCGACGACGACCCGATGCAATGTCGACGCGGCGCTGGCG 2191573

```

```

231 lUserMetLysGluThrLeuArgIleArg.....ArgLeuGlyGlu 244
|||||
2191574CGGGGGGTGCCGCTCGATGTCGCCGCTGACGAGCGCGCGATGGCGCA 2191623

```

```

245 ProGluAspCysAlaGlyIleValSerPheLeuGlySerGluAspAla 261
|||||
2191624ACCGAGAGATGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191673

```

```

261 rTyrlleThrGlyLysThrValValAlaGlyGlyThrProSerArgL 278
|||||
2191674GATGATCAGCGCGCACCCAGATCGCGCGGAGCGGACGATTCGCCGC 2191723

```

```

278 eu 278
||
2191724TG 2191725

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:

```

; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

alignment_scores:

```

Quality: 368.00      length: 284
Ratio: 1.957         Gaps: 8
Percent Similarity: 66.197      Percent Identity: 35.563

```

alignment_block:

US-09-866-034-2 x US-09-103-840A-1

Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529

```

10 CysAlaArgAlaTRP.....AsnSerValAr 18
|||||
2193596TGTGCGCAGATTGGGCTTGCATGTTCCGGCCGATCCGAGAACCCGACGA 2193645
18 gMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysVala 35
|||||
2193646CGTTGGCGCGCCGAGGTTGAAACCATCCAGAC...CTGGCGCGGCAAGGTGC 2193692
35 LaleuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 51
|||||
2193693CGATCGTTGGGCGCGCGCGGCGGGAATCGCTGGCGGTTGCCCGCGCA 2193742
52 LeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGlnGlnAs 68
|||||
2193743CTGCGCCGACGAGGCTGCCATGTGCTGCGCGGACATCGATGGTGATGC 2193792
68 nValAspGlnAlaValAlaThrLeuGlnGlyGluGlyLeuSerValThrG 85
|||||
2193793CGCGGAT...GCCGCGCCACCAAAATCGGTGTGCG...G 2193827
85 LyrThrValCysHisValGlyLysAlaGluAspArgGlu.....ArgLeu 99
|||||
2193828CAGCGCGCTGCCGCTTACGTCACGACGACACACACATCATCGCCATG 2193877
100 ValAlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAs 116
|||||
2193878GTGACGCCCTGTTGCCCGCTTCCGCGGCGGAGTGGACAACTGGTGCACA 2193927
116 nAlaAlaValAsnProPhePheGlySerIleMetAspValThrGluGlu 133
|||||
2193928CGCGCGCTGTC...GTTCATCTGCTGCTGCTCATGACACACCGTCGAG 2193974
133 alTrpAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThr 149
|||||
2193975ACTTCATGCGGATCGCATCGCATCATCTCCGCGCGCTGGCTGCATCC 2194024
150 LysAlaValAlaValProGluMetGluLysArgGlyGlyIleGlySerVal 166
|||||
2194025AAGCATGCGGACCGCGCATGATCGAGCGCGCGGCGGACCCATTGTCA 2194074
166 eValSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyr 183
|||||
2194075CCTGTCGTAGCGGCGCAGGTACGCGTGGCGGACCGCGGCATPAG 2194124
183 snValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 199
|||||
2194125GCATGTCGAAAGCGCGCATCATCCAGCTCAGCCGATCACCGCGCGCA 2194174
200 LeuAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIle 216
|||||
2194175CTGCGCTGCTCGCGCATCGCTCCAAACGCTGCTCCCGCATTCGTCA 2194224
216 sThrSerPheSerArg.....MetLeuTrpMetAspLysGluLysGlu 231
|||||
2194225CACCCGATGACAGACCGCATGTCGATGTCGACGCGGCGCGCTGGG 2194274
231 LysMetLysGluThrLeuArgGlyLeu.....ArgLeuGlyGlu 244
|||||
2194275CGGCGGCTCGCGCATGATGTCGCGCGCTCGAGCGCGCATGCGCGCA 2194324
245 ProGluAspCysAlaGlyIleValSerPheLysSerGluAspLase 261
|||||
2194325CCGAGAGAAATGCGCGCATGCTGCTGCTGCTGCTGCGAGCATCGCTC 2194374
261 rTyrIleThrGlyGluThrValAlaValGlyLysGlyThrProSerArg 278
|||||
2194375GATGATCACCGCGCACCCAGATCGCGGAGGCGGAGATTCGCGCC 2194424
278 eu 278
||

```

2194425TC 2194426

seq_name: /cgn2_6/plodata/2/ina/6b_COMB.seq:US-09-468-738A-28

```

seq_documentation_block:
; Sequence 28, Application US/09468738A
; Patent No. 6312933
; GENERAL INFORMATION:
; APPLICANT: kimoto, No. 6312933,hiro
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Mitsuhashi, Kazuya
; TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, D
; TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING
; FILE REFERENCE: 06501-050001
; CURRENT APPLICATION NUMBER: US/09468,738A
; CURRENT FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: JP 1999-171160
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: JP 1998-363130
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0, reformatted using wordperfect 5.1
; SEQ ID NO: 28
; LENGTH: 786
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-468-738A-28

```

alignment_scores: Quality: 331.50 Length: 250
Ratio: 1.973 Gaps: 6
Percent Similarity: 67.200 Percent Identity: 34.000

alignment_block: US-09-866-034-2 x US-09-468-738A-28

Align seg 1/1 to: US-09-468-738A-28 from: 1 to: 786

```

30 LeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGlyIleGly 46
|||||
13 TTTAAAGAAAGAACTCGCTATTACAGAGCTGCTTACAGCGCTCGGAA 62
46 eAlaIleAlaArgArgLeuAlaGlnAspGlyAlaHisValAlaValSer 62
|||||
63 GCGGATGCGCATTCGCTTGGCAAGAGCAAGCAAAAGTGTATCACT 112
63 .....SerArgLysGln.....GlnAsnValAspGlnAlaValAlaThr 75
|||||
113 ATTATGATTAATAACAAGATCCGACGAGGTAATAAGAGGTCAATCAG 162
76 LeuGlnGlyGluGlyLeuSerValThrGlyThrValCysHisValGly 92
|||||
163 GCGGCGGTGAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 203
92 sAlaGluAspArgGluArgLeuValAlaThrAlaValLysLeuHisGly 109
|||||
204 AGAGAAAGATGTAATAAATATCTGCAAAACGCGCAATTAAGGACTGCGCA 253
109 LysIleAspIleLeuValSerAsnAlaVal.....AsnProPhePheGly 124
|||||
254 CACTGCATATATGATTAATAATGCGCGCTTGAATAATCTGTGCCATCT 303
125 SerIleMetAspValThrGluGluValThrAspLysThrLeuAspLase 141
|||||
304 CACGAATATCGCTCAAGCAT...TGGATTAAGTCAATCGCGCA 347
141 nValLysAlaProAlaLeuMetThrLys...AlaValProGluMetG 157
|||||
348 CTTAAGGCTGCTTTTGAAGCGCGTGAAGCATTAATATTTGCTAG 397
157 LysArgGlyGlyLysSerValAlaIleValSerIleAlaAlaPhe 173

```

seq_documentation_block:

LENGTH: 1155 base pairs

US-08-793-035-2

Quality: 330.50

Percent Similarity: 67.600

US-09-866-034-2 x US-08-793

Align seg 1/1 to: US-08-79

232 GTGGTCGTTGTGACTGGTG

50 gargLeuAlAGlnAspGLY

282 TTCCTGGGCAAGCTGGC

63 erArgLySGInGInAsnVa
.....

332 CAAGGAGGCTGAGGAAGT

80 GlyLeuSerValThrGlyT
.....|||

382 GCTATTACTTTTGGGGGT.

96 gGIuArgLeuValAlaThr
||| |||

423 GGAAGCCATGATGAAAACC

```
113 euValSerAsnAlaLaVa
..|||.....|.....
```

473 TCGTCAACAATGCAGGAAT

130 TrnGLuValTrpAspL
.....|

520 AAGGAGTCCCAATGGGATG

146 aleumetInrlySALaval

5' U TCTCTGTACCCAGGCAGCA

```
163 evalValIleValSerSe
    .....4.....
```

620 GAATCATCAACATTGCCGTC,

180 SerProTyrAsnValSerLj

6 / 0 GCAACCTACGCTGCTGCTA

196 UALALEGIULEUALA PRO

120 CCGCAGAGAGGGTTCGAGGCA

213 LysLeuLysThrSerPhe
||::|||
.....

GGTTCATTGCAATCTGACATC

230 6UGUSERMETLYSG1UTH

820 AAA.....ATCTTGGGAAC

240 uaspcysalaclylleval

864 AGA TGTGGCTTGGCTTGGTAC

262 YLLIEIHRGLYGLUThrVal

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-793-035-1

seq_documentation_block:

Sequence 1, Application US/08793035

Patent No. 6011201

GENERAL INFORMATION:

Applicant: Slabas, Antoni R.

Applicant: White, Andrew

Applicant: Chase, Dianne

Applicant: Elborough, Keiran

Applicant: Fentem, Philip A.

Title of Invention: B-ketocycl ACP Reductase Genes From

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: US

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/793.035

FILING DATE: 28-JUL-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9414622.2

FILING DATE: 20-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB95/01678

FILING DATE: 17-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kammerer, Patricia A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: MOBT-132

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713.787.1400

TELEFAX: 713.787.1440

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1185 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-793-035-1

Alignment scores:

Quality: 330.50

Ratio: 1.956

Percent Similarity: 67.600

Percent Identity: 33.600

Alignment block:

US-09-866-034-2 x US-08-793-035-1

Align seg 1/1 to: US-08-793-035-1 from: 1 to: 1185

```

34 ValAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaR 50
|||||
289 GTGTCGCTGTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 338
50 gAArgLeuAlaGlnAspGlyAlaHisValValValValValValValVal 63
|||||
339 TTCCTTGAGCAAGCTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 388

```

```

63 eArGlySGlnGlnAsnValAspGlnAlaValAlaThrLeuGlnGly 79
|||||
389 CAAAGAGGCTGAGAGAGCTTCTTAACAGATTGACAGCTATGAGGCCAG 438
80 GlyLeuSerValThrGlyThrValCysHisValGlyLysAlaGlnAspAr 96
|||||
439 GCTATCTACTTTGGGGCT.....GATGCTCCCAAGAGGCTGATGT 479
96 gGUArGLeuValAlaThrAlaValLysLeuHisGlyGlyIleAspIle 113
|||||
480 GGAAGCCATGATGAAACCCGCTTGTGTCGTCGTCGTCGTCGTCGTCG 529
113 euValSerAsnAlaAlaValAsnProPheGlySerIleMetAspVal 129
|||||
530 TCGCACAATGACGAGGATCAGCTCG...CATACCTTGTCGATGACGATG 576
130 ThrGluGluValTrpAspLysThrLeuAspIleAsnValLysAlaPro 146
|||||
577 AAGAAGTCCCAATGGATGAAAGTATGATGATGATGATGATGATGATG 626
146 aleuMetThrLysAlaValAlaValProGluMetGluLysArgGlyGly 163
|||||
627 TCTCTGTACCCAGGACAGCAAAAGATCATGATGAAAGCAAGAGGAA 676
163 eValValIleValSerSerIleAlaAlaPheSerProSerProGlyPhe 179
|||||
677 GATTCATCAACATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
180 SerProTyraAsnValSerLysThrAlaLeuGlyLysLeuThrLysThr 196
|||||
727 GCAAACTACGCTGCTCTAAAGCTGCTGATGATGATGATGATGATGATG 776
196 uAlaIleGluLeuAlaProArgAsnIleArgValAsnGlyLysAlaPro 213
|||||
777 CGCAGAGAGGTCGCGAGCAATGAAATGCAATGCTGCTGCTGCTGCTG 826
213 lyeuIleLysThrSerPheSerArgMetLeuTrpMetAspLysGlyLys 229
|||||
827 GCTTCATTCGATGTCAGATGACGTCGTCGTCGTCGTCGTCGTCGTCG 876
229 GluGluSerMetLysGluThrLeuArgIleArgArgLeuGlyLysGluPro 246
|||||
877 AAA.....ATCTTGGGACAAATCCATTCATGACGATTCGTCGTCG 920
246 uAspCysAlaGlyIleValSerPheLeuGlyLys...SerGlnAspAlaSer 262
|||||
921 AGATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
262 yTleThrGlyGluThrValValValValValValValValValValVal 277
|||||
971 ACATCAGAGAGGCTGTCACCATTCATGAGAGGATTTGTCATCTAGG 1018

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-363-189B-5

seq_documentation_block:

Sequence 5, Application US/09363189B

Patent No. 6242228

GENERAL INFORMATION:

Applicant: SUGIYAMA, MASAKAZU

Applicant: TONOUCHI, NAOTO

Applicant: SUZUKI, SHUNICHI

Applicant: YOKOZAKI, KENZO

Title of Invention: XYLITOL, DEHYDROGENASE OF ACETIC ACID BACTERIA AND GENE THERE

FILE REFERENCE: 0010-1024-0

CURRENT FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: JP10-216047

PRIOR FILING DATE: 1998-07-30

NUMBER OF SEQ ID NOS: 16

SOFTWARE: Patentin version 3.0

SEQ ID NO 5

LENGTH: 2774

TYPE: DNA

171 .CTGGAAGGGAAGCT

```
46 heaiallaaiaargleuAlaaspGlyAlaHisValValSer 62
   |||||
220 AGGCATCTGAGGCTGTTAAAGCAGCGGCCAAGGTGATCGCG 269
63 SerArgLysGlnGlnAsnValaspGlnAlaValAlaThrLeuGlnGly 79
   |||||
270 .....GACATCGACGCGCGGCG.....GGCGA 292
79 uGlyLeuSerVal.....ThrGlyThrVal...CysHisVal 90
   |||||
293 GCGCGTGGCGGCGGCGCTGGCGCCGACGCTCGGCTTCGCGGTGGAGC 342
90 aGlyLysAlaGlnAspArgLysValValAlaThrAlaValLysLeu 106
   |||||
343 TGTCTGTGGAGGAGCAGTGTGAGCGCGCTGAGCGCGCGCTGGCGG 392
107 HisGlyGlyIleAspIleLeuValSerAsnAlaAlaValAspProPhe 123
   |||||
393 TAGCGGCGCTGAGCGTGTGCAACAACGCGCGGTG.....CT 433
123 eGly.....SerIleMetAspValThrGlnGlyVal 133
   |||||
434 GCGCGCCGACAGCGCGCGCCCAAGCATCTGTCGTCGACGCGCGG 483
133 aITrpAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThr 149
   |||||
484 AGTTCGACCGCGTCTCGCGCTCAACGCGCTGGCGCGCGCTGGCAG 533
150 LysAlaValValProGlnMetGlnLysArgLysGlySerValValI 166
   |||||
534 AAGCAGCGGCGCGCTGCGCATGACCGCGCGCGCGCGCGCATCATC 583
166 eValSerSerIleAlaAlaPheSerProGlyPheSerProTyrA 183
   |||||
584 CGTCCGACAGCTGCGCGCGCTGCTGCGGCGCTCGCGCGCGCGCTTCA 633
183 snValSerLysThrAlaLeuGlnGlyLeuThrLysThrLeuAlaIleG 199
   |||||
634 CCGCTCCAGCAGCAGCATCGTGGCGCTCACCAACAGCGCGCTGGCAG 683
200 LeuAlaProAlaArgAsnIleArgValAsnGlyLeuAlaProGlyLeuI 216
   |||||
684 CTCGGCGCGCGCGCATCGCGCTCACTGATCTCCCTTGGCGGTGCG 733
216 sTherPheSerArgMetLeuTrpMet..... 225
   |||||
734 CACCGCGATGCTCATCAACGCGCTGGCGCACGACGCGCTCACCG 783
226 .....AspLys 227
784 CCGACGACGCGCGACGCGCATCGACCTCGACATGCGCGCTGCGCGAC 833
228 sLysGlnGlnGlnSerMetLysGlnThrLeuArg..... 238
   |||||
834 CAGGAGGTGAGAAATGATGAGAGGTGTGAGGCGCTCGCGACGCTCAA 883
239 .....IleArgArgLeuGlnGlnProGlnAspCysAlaGlyIleVal 252
   |||||
884 GCGCGCGCGCTGAGA.....CCGAGGACATCGCGCGGAGGCGG 921
252 aLserPheLeuGlnSerGlnAspAlaSerTyrIleThrGlyLysVal 268
   |||||
922 CGCTCTTCTCGCGCAGCAGCATCTCAATATTCGCGCGCACACCTC 971
269 ValValGlyGlyGly...ThrProSerArg 277
   |||||
972 GTCGTGACGCGCGCGCTGACACCTCCAGA 1001
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-241-766-12
seq_documentation_block:
; Sequence 12, Application US/08241766
; Patent No. 568590
```

```

GENERAL INFORMATION:
APPLICANT: JACOBS, W. R.
APPLICANT: COLLINS, D. M.
APPLICANT: BANERJEE, A.
APPLICANT: DELISLE, G. W.
APPLICANT: WILSON, T. M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,766
FILING DATE: 12-MAY-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20003.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 494..1234
FEATURE:
NAME/KEY: CDS
LOCATION: 1256..2062
US-08-241-766-12
alignment_scores:
Quality: 321.00 Length: 248
Ratio: 1.922 Gaps: 5
Percent Similarity: 67.339 Percent Identity: 31.855
alignment_block:
US-09-866-034-2 x US-08-241-766-12
Align seg 1/1 to: US-08-241-766-12 from: 1 to: 2232
27 ArgAspProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspG 43
   |||||
521 AAACCCCATTCGTATCCGTCAGTCTCGGTACGAGGAAACCGGGG 570
43 yLleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAlaHisVal 60
   |||||
571 GATCGGGGTGGCGATCGCAGCGGCTGCTCCGACGCGCCCAAGGTGG 620
60 aValSerSerArgLysGlnGlnAsnValaspGlnAlaValAlaThrLeu 76
   |||||
621 CCGTCACGACCGGTGA..... 637
77 GlnGlyGlnGlyLeuSerValThrGlyThrValCysHisValGlyLysAl 93
   |||||
638 TCCGAGGCGCCAAAGGCGCTGTTGGCGTGAATGTGACGTCACGACGAG 687
```

```

93 agluaspargluargluvalalathralaallysleuhs...glyc 109
688 CGAGCGCGTCGATGCG...GCCCTCAGCGCGCTAGAGACGACCGAGGTC 734
109 lytleaspilleuvalserasnalaalavalasnprophephglyser 125
735 CGGTCGAGGTCGTGTGTCCACCGCGGCTATCCGGC...GAGCATTC 781
126 llemetaspvalthrugluvalatrpaspysrthrleuaspilleasna 142
782 CTCATGCGGATGACGAGGAAAGTTCGAGAGTCATCAACGCCAACCT 831
142 llvalspvalaleumethrlysalavalaproglumetglulysa 159
832 CACCGGGCGTCCGGGTCTCAACCGGCATCCGCGCATGACGCGCA 881
159 rglglyglyservalvalillealserlelealaphserpro 175
882 ACAATTCGGTCGATGATATTCATAGGTGCTCCGCGAGCTGGGCGC 931
176 Serproglyserprotyrasnvalserlysrthralaleuaglyle 192
932 ATCGGCAACGAGCGACACTACGAGCGCTCCAGCGCGAGTGGCAT 981
192 uthrlysrthleualalleuvalaalaalproargasnilleargvalasnc 209
982 GCGCGCGCTCGATGCGCGCGCTGTCGAGCGCAACGTCGACCGCATG 1031
209 ysleualaprogllyleuilelysrthserpheserargmetleutrmec 225
1032 TGCTGGCGCGCGGCTACATGACACGATATGACCGCGGCTC... 1075
226 AsplysgluylsgluuasermetlysgluThrleuarglleargyle 242
1076 GATGAGCGGATTCAGCAGGCGCGCTCATTTATCCACGAGAGCGGCT 1125
242 usglyuaprogluaspcysalaelylevalserpheleucysserglua 259
1126 GCGGACCCCGCGGAGGTCCGCGGGGTGTCACGCTTCGCTGCTCCGAG 1175
259 spalaserlyrlethrlygluThrvalavalaglygly 273
1176 ATCGAGCTATATCTCCGCTGCGGCTATCCCGGTGCGAGCGGCGC 1219

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-491-146-1
seq_documentation_block:

```

```

: Sequence 1, Application US/08491146
: Patent No. 5556778

```

```

: GENERAL INFORMATION:
: APPLICANT: Sacchetti et al
: TITLE OF INVENTION: INHA CRYSTALS AND THREE
: TITLE OF INVENTION: DIMENSIONAL STRUCTURE
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESS: Amster, Rothstein & Edeinstein
: STREET: 90 Park Avenue
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10016

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
: COMPUTER: IBM PC Compatible

```

```

: OPERATING SYSTEM: MS-DOS

```

```

: SOFTWARE: Word Processor (ASCII)

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/491.146

```

```

: FILING DATE:

```

```

: CLASSIFICATION: 435

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/307.376

```

```

: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: George, Kenneth P.
: REGISTRATION NUMBER: 30,259
: REFERENCE/DOCKET NUMBER: 96700/296
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5995
: TELEFAX: (212) 286-0854 or 286-0082
: TELEX: TWX 710-581-4766
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3120
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: both
: MOLECULE TYPE: DNA
: DESCRIPTION:
: HYPOTHETICAL: NO
: ANTI-SENSE:
: FRAGMENT TYPE:
: ORIGINAL SOURCE: inhA operon
: ORGANISM: M tuberculosis
: STRAIN:
: INDIVIDUAL ISOLATE:
: DEVELOPMENTAL STAGE:
: HAPLOTYPE:
: TISSUE TYPE:
: CELL TYPE:
: CELL-LINE:
: ORGANELLE:
: IMMEDIATE SOURCE: M tuberculosis
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT:
: MAP POSITION:
: UNITS:
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: PUBLICATION INFORMATION: No. 5556778e
: AUTHORS:
: TITLE:
: JOURNAL:
: VOLUME:
: PAGES:
: DATE:
: DOCUMENT NUMBER:
: FILING DATE:
: PUBLICATION DATE:
: RELEVANT RESIDUES IN SEQ ID NO:
: US-08-491-146-1

```

```

: alignment_scores:

```

```

: Quality: 321.00 Length: 248
: Ratio: 1.922 Gaps: 5
: Percent Similarity: 67.339 Percent Identity: 31.855

```

```

: alignment_block:

```

```

: US-09-866-034-2 x US-08-491-146-1

```

```

: Align seg 1/1 to: US-08-491-146-1 from: 1 to: 3120

```

```

27 ArgaspProleuAlaAsnlysalalaleuValThrAlaserThrAsp1 43
: : : : : : : : : : : : : : : : : : : : : : : : : : :
251 AAGCCCAATTCGTATCCGCTTCAGTCCCTGTTACCGGAGAAACGGGG 300
43 ylleglyrleuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 60
: : : : : : : : : : : : : : : : : : : : : : : : : : :
301 GATCGGCGTGGCGATGCGACAGCGGCTGCGCGGCGGCGGCGGCGG 350
60 alValserSerArglysgInglInAsnValAspAlaAlaAlaThrLeu 76

```

```
351 CCCTCACCCACCGCTGA..... 367
      ::|||::: |||
77 gln glyglu glyleu serValThr glyThrValCysHisValGlyLysAl 93
      ::|||::: |||::: |||::: |||
368 TCCGGAGGCCCAAGGGCGCTGTTGGCGTCGAATGTGACGTACCCGACAG 417
      ::|||::: |||::: |||::: |||
93 agl u s p a r g g l u a r g l e u v a l a l a t h r a l a v a l y s l e u H i s . . . G l y c 109
      ::|||::: |||::: |||::: |||
418 CGACGCCGTCGATCCG...GCCCTCACGGCGGTGACAGACACACGAGGTC 464
      ::|||::: |||::: |||::: |||
109 l y l e a s p l e u v a l s e r a s n a l a l a v a l a s n p r o p h e g l y s e r 125
      ::|||::: |||::: |||::: |||
465 CGGTGAGGTGCTGGTGTCCAAAGCCGCGCTATCCGCG...GACGCATTC 511
      ::|||::: |||::: |||::: |||
126 l l e m e l a s p v a l t h r g l u g l u v a l t r p a s p l y t h r l e u a s p l l e a s n v a 142
      ::|||::: |||::: |||::: |||
512 CTCATCCGATGACCCGACGAGAAAGTTCCGAGAAAGGTGATCAACGCCCAACT 561
      ::|||::: |||::: |||::: |||
142 l l y s a l a p r o a l a l e u m e t t h r l y s a l a v a l p r o g l u m e t g l u l y s a 159
      ::|||::: |||::: |||::: |||
562 CACCGGGCGCTCCGGGTGCTCAAGCGGCAATCGCGCAGCATGACAGCGCA 611
      ::|||::: |||::: |||::: |||
159 r g l y g l y s e r v a l v a l l e v a l s e r s e r l l e a l a l a p h e s e r p r o 175
      ::|||::: |||::: |||::: |||
612 ACAAAATCGGTGAATGATATTCATAGTTCGGTCTCCGCGACCTGGGGC 661
      ::|||::: |||::: |||::: |||
176 s e r p r o g l y p h e s e r p r o t y r a s n v a l s e r l y s t h r a l a l e u l e u g l y l e 192
      ::|||::: |||::: |||::: |||
662 ATCGGCACACGACGACACTACGACGACCTCCAGAGCCGAGGTGATGCAAT 711
      ::|||::: |||::: |||::: |||
192 u t h r l y s t h r l e u a l a l l e g l u l e u a l p r o a r g a s n l e a r g v a l a s n c 209
      ::|||::: |||::: |||::: |||
712 GGCCTCGCTGATCGCCCGGAGCTGTGAGAGCAAACTGACCGCAATG 761
      ::|||::: |||::: |||::: |||
209 y s l e u a l a p r o g l y l e u l l e l y s t h r s e r p h e s e r a r g m e t l e u t r p m e t 225
      ::|||::: |||::: |||::: |||
762 TGGTGGCCCCGGGCTACATCGACACCGATATGACCCGCGCTG..... 805
      ::|||::: |||::: |||::: |||
226 A s p l y s g l u l y s g l u g l u s e r m e t l y s g l u t h r l e u a r g l l e a r g a r g l e 242
      ::|||::: |||::: |||::: |||
806 GATGAGCGGATTCACGAGGGGCGCTGCATATTATCCACGAGAGCGGCT 855
      ::|||::: |||::: |||::: |||
242 u g l y l u p r o g l u a s p c y s a l a g l y l l e v a l s e r p h e l e u c y s s e r g l u a 259
      ::|||::: |||::: |||::: |||
856 CGGCACCCCGCGAGGTGCGCGGGGTGCTGCTGCGCTCCGAGG 905
      ::|||::: |||::: |||::: |||
259 s p a l a s e r t y r l l e p h r g l y g l u t h r v a l v a l a l g l y g l y 273
      ::|||::: |||::: |||::: |||
906 ATGCGAGCTATATCTCGGTGCGGTATCCGCTGACGCGCGG 949
```


seq_name:	gb_hfc:BC000663		
seq_documentation_block:			
LOCUS	BC000663	1298 bp	mRNA
DEFINITION	Homo sapiens, Similar to peroxisomal short-chain alcohol		Linear
			HTC 12-JUL-2001

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC), Bethesda, MD

Web site: <http://www.nisc.nih.gov/nisc-mcgenbri.nih.gov>
Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamini, B., Blakesley, R.W., Houtford, G.G., Brinkley, C., Dietrich, N.L., Gunn, X., Gupta, J., Ho, S.-I., Karlins, E., Legaspil, R., Lim, M., Maduro, Q.L., Mastaglio, C., Mastriani, D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stanitrop, S., Thomas, P.J., Tlionson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: <http://image.liml.gov>
 Series: IRAL Plate: 5 Row: d Column: 12
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 4105189
 This clone has the following problem: frame shifted.

FEATURES	Location/Qualifiers
source	1. .1298

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:33493226"
/tissue_type="Colon, adenoc
/clone_1ib="NIH_MGC.15"
/adb_host="DH10B-R"
/notes="Vector: pOTB7"
BASE COUNT      311 a      259 t
ORIGIN           388 c      400 g

```

```
alignment_scores:
|
```

Quality:	1289.00	Length:	278
Ratio:	4.756	Gaps:	0
Percent Similarity:	97.482	Percent Identity:	95.324

```
alignment_block:
```

Align seg 1/1 to: BC000663 from: 1 to: 1298

1 MethistylsylaGlyLeuLengIYleucysAlaargAlatrpanserVa 17
 ||||| ||| ||| |||
 48 ATGCACATAGGCCAGGCTCTAGGCCCTTGTCCTCGGGCACGGAATCCGT 97
 |||
 17 LargMetalaserseirgIyMetThrArgARASPProleualasnlvsV 34
 |||||:::||||| |||
 98 GCGGTTGGCCAGCTCCAGAGATTACCGCGGGGACCCTCACAATTAAG 14
 |||
 34 alAlaleuValThrlaserThraspcIyIllegllyphealIealarg 50

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: gcaps-remail.nih.gov
           Tissue Procurement: DCTD/DTF
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA sequencing by: Incyte Genomics, Inc.
           <Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
           Plate: L1CWB3      row: c      column: 08
           High quality sequence start: 20
           High quality sequence stop: 782.
           Location/Qualifiers
             1..782
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3937471"
               /clone_1id="N1H.MGC.7"
               /tissue_type="small cell carcinoma"
               /cell_line="MGC3"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: Lung; Vector: pOTB1; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally
               cloned into EcoRI/XhoI sites using the following 5'
               adaptor: GGCACGAC(G). Size-selected >500bp for average
               insert size 1.8kb. Library constructed by Ling Hong in
               the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      172 a      200 c      263 g      147 t
ORIGIN

alignment_scores:
  Quality: 1257.00      Length: 251
  Ratio: 5.008      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-866-034-2 x BE798992

Align seg 1/1 to: BE798992 from: 1 to: 782

3      LysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerValArgme 19
30      AAGGGGGGGGCTGCTAGAGCCCTCTGTGCCCGGCGCTTGAAATTCGGCGGAT 79
19      LAlaSerSerGlyMetThrArgAlaGAspProLeuAlaAsnLysValAlaL 36
80      GGGCAGCTCCGGGATGACCCCGGGAGACCCCTGCAATATAGGTGGGCC 129
36      euValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgLeu 52
130      TGGTAACGGCTCCACCGACGAGATCGGCTTCGCCATGCGCCGCGTTTG 179
53      AAlaGlnAspGlyAlaHisValValAlaSerSerArgLysGlnAsnVa 69
180      GCCCAGGACGGGGCCCATGTGTCAGCAGCGGAGGACGAGCAATGT 229
69      LAspGlnAlaValAlaThrLeuGlnGlyGlyGlyLeuSerValThrGlyT 86
230      GGACGAGCGGTGGCCACGCTGCAGGGGGAGGGGCTGAGCGTGAGGGCA 279
86      hrValCysHisValGlyLysAlaLeuAspArgGlnArgLeuValAlaThr 102
280      CCGTGTGCATGTGGGAGGCGGAGGACCGGAGCGGCGGTGTGTGCCACG 329
103      AlAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAlaAla 119
330      GCTTGTGAAGCTTCATGAGGATATCCATATCTCCATATGCTGCTGT 379
119      LAspProPheGlySerIleMetAspValIleThrClnGlnValThrTran 176

```

```

|||||
380 CAACCCCTTCTTGAAGCATATGATGTCACTGAGGAGCTGGGACA 429
136 ysthrluaspilleasvalysalaproalaueuthrllysalaal 152
430 AGACTCTGGACATTATGTCAAGGCCCAAGCCCTGATGACAAAGCAGT 479
153 ValProgluMetGluysArgylgylgylserValValleValser 169
480 GTGCCAGAAATGAGAAACAGAGGCGGCTCACTGTGTCGTCTTC 529
169 rlealaalapheserproserproglpgheserProTyAsnValser 186
530 CATAGCAGCCTTCAGTCATCTCTGCTTCACTCTTCAATGTCAGTA 579
186 ysthrlaaleuenglyleuthrllysThrlleuallelelualeapro 202
580 AAACAGCCTTGTCTGGCCCTGACCAAGACCTGCGCATAGAGCTGGCCCA 629
203 ArgAsnleArArgValAsncysleualaproglyleuileysThrsrph 219
630 AGGACACTTGGGTGAACCTGCTGACACCTGACTTATCAAGACTGACTT 679
219 eSerArgMetleuTrpMetAspIlysgIuysgluIuserMetLysglut 236
680 CACGAGATCTCTGGATGACAAAGAAAAGAGAAAGCATGAAAGAA 729
236 hrleuarglleargargleuglygluprogluaspysalaglileal 252
730 CCTGGGATTAAGAGATTAGGCGACGACGAGGATTTGCTGTCATCTG 779
253 Ser 253
780 TCT 782

seq_name: gb_est2:BI334186

seq_documentation_block:
LOCUS      BI334186                819 bp      mRNA      linear      EST 30-JUL-2001
DEFINITION 602997649P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5139733 5',
            mRNA sequence.
ACCESSION  BI334186
VERSION    BI334186.1  GI:15018843
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 819)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/MLN at:
            http://image.llnl.gov
            Plate: LHM11344  row: a  column: 14
            High quality sequence stop: 817.
            Location/Qualifiers
                1..819
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5139733"
                /clone_lib="NIH_MGC_12"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /note="Organ: cervix; Vector: PCMV-SPORE6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

```

```

Average insert size 1.4 kb. Library prepared by Life
Technologies.
BASE COUNT      186 a      211 c      265 g      157 t
ORIGIN
alignment_scores:
    Quality: 1209.50
    Ratio: 4.762
    Percent Similarity: 95.131
    Percent Identity: 93.633
alignment_block:
US-09-866-034-2 x BI334186
Align seg 1/1 to: BI334186 from: 1 to: 819
1  MethIstysAlaGlyLeuEnGlyLeuCyAlaArgAlaTrpAsnSerVa 17
||||| ||| ||||||||||||||| ||| |||
23  ATGCACATGSCCAGGCTGCTAGGCTTGTGCTTGGCCAGCAGGAGTCGT 72
17  lArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
73  GCGGATGGCCAGCTCCAGATGACCCCGCGGAGCCGCTCACAAATAGG 122
34  alAlaLeuValThrAlaSerThrAspGlylleGlyPheAlaIleAlaArg 50
123  TGGCCCTGTAAACGGCTCCACGACGAGTCGCGCTTGCATCGCCCGG 172
51  ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgLysGln 67
173  CGTTTGGCCCAAGGACAGAGGCCCACTGCTGCTACGACGCGAGAGCA 222
67  nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValTr 84
223  GAATGTGACACAGCGGGTGGCCACCGCTGCAGGGGAGAGGCTGAGCTGA 272
84  hrGlyThrValLysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
273  CGGGCACTGTGTGCCATGTGGGAGGCGGAGGACCGGAGCGGCTGCTG 322
101  AlaThrAlaValLysLeuHisGlyGlyLysAlaSerLysValSerAsnAl 117
323  GCCATGCTGTGAAGCTTTCATGAGGATTCATATCTCTTCCAAATGC 372
117  aaAlaValAsnProPhePheGlySerIleMetAspValThrGluGluValTr 134
373  TGTGTTCACACCTTCTTGTGGAGCCTAATGATGTACCGAGAGAGGTGT 422
134  rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
423  GGGACAAGACTCTGACATTAATGTGAAGGCCCAAGCCCTGATGACAAAG 472
151  AlaValValProgluMetGluysArgylgylgylserValValleVal 167
473  GCAGTGTGTCAGAAATGAGAAACAGAGGCGGCTCACTGTGATGATCCT 522
167  lSerSerIleAlaAlaPheSerProserProglpgheserProTyAsnV 184
523  GTTTCATATGACAGCTTCACTGCTCACTCTGCTTCACTTACATG 572
184  aLserLysThrAlaLeuEnGlyLeuThrLysThrleuAlaIleGluLe 200
573  TCAGTAAACAGCCTTCTGCTGCGCTGACCAAGACCTGCGCATGAGA...GC 619
200  uAlaProAlaArgAsnleArArgValAsncysleualaproglyleuile 217
620  TGGCCCAAGGAAACATTAGGTGAACCTCCATGACCTGCACTTATCAAGA 669
217  hrSerPheSerArgMetleuTrpMetAspIlysgIuysgluIuserMet 233
670  CTAGCTTCAGCAGATGCTGTGATGACAAAGAAAAGAGAAAGCATG 719
234  LysGluThrleuArglleargargleuglygluprogluaspysalagl 250

```


Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM523 row: 0 column: 16
 High quality sequence start: 13
 High quality sequence stop: 779
 Location/Qualifiers
 1. 974
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3837927"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 211 a 243 c 334 g 186 t
 ORIGIN

alignment_scores:
 Quality: 1198.50 Length: 270
 Ratio: 4.663 Gaps: 6
 Percent Similarity: 95.185 Percent Identity: 94.074

alignment_block:
 US-09-866-034-2 x BE744486

Align seg 1/1 to: BE744486 from: 1 to: 974

```

8 GlyLeuCySAlaArGaLaTrPaNsSerValArGMeLaSerSerGlyMe 24
|||||
10 GGCCCTGTGCTCCGGCTTGGAAATTCGGTCCGATGCCAGCTCCGGGAT 59
|||||
24 tTrrArgArGAspProLeuAlaSnLysValAlaLeuValThAlaSerT 41
|||||
60 GACCCCGGGGACCCGCTCCAAATAGAGTGGCCCTGGTAAACGGCTCCA 109
|||||
41 hrAspGlyIleGlyPheAlaIleAlaArGArGLeuAlaGAspGlyAla 57
|||||
110 CCGACGGGATCGGCTTCCCATGCGCCGGGCTTTGGCCCGGAGGCGCC 159
|||||
58 HisValAlaValSerSerArGlySGInGInAsnValAlaSpGInAlaValAl 74
|||||
160 CATGTGCTGCTCAGACCGGAGACGAGCAATGTGACACGCGGCGTGGC 209
|||||
74 aTrhLeuGInGlyGlyLeuSerValThrGlyThrValCysHisValG 91
|||||
210 CAGCGTCGAGGGGAGGGGCTGAGCGTGAACGCGACCGTGGCATGTGG 259
|||||
91 lYlSAlaGluAspArgGluArgLeuValAlaThrAlaValLysLeuHis 107
|||||
260 GGAAGCGGAGGAGCGGAGCGGCTGTGGCCACGGCTGTGAAGCTTCAT 309
|||||
108 GlyGlyIleAspIleLeuValSerAsnAlaAlaValAsnProPheG 124
|||||
310 GGAGGTATCGATTCCTAGTCTCCATGCTGCTCAACCTTTCTTTGG 359
|||||
124 ySerIleMeLaSpValThrGluGluValTrpAspLysThrLeuAspIleA 141
|||||
360 AACCAATAAAGATGCTCAGAGAGGTGTGGACAAGACTCTGCACATTA 409
|||||
141 snValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetGlu 157
|||||

```

```

410 ATGTGAAGGCCCAAGCCCTGATGACAAAGCATGTGTGCCAGAAATGAG 459
158 LysArgGlyGlyGlySerValValIleValSerSerIleAlaAlaPhe 174
|||||
460 AAAGGAGAGGCGGCTCGATGCTGATGCTCTTCCATAGAGGCTTCAG 509
|||||
174 rProSerProGlyPheSerProTyraSnaValSerLysThrAlaLeuG 191
|||||
510 TCATATTCCTGCTTCACTGCTTCAATGTGATTAACACCTTGTGCTGG 559
|||||
191 lYleuThrLysThrLeuAlaIleGluLeuAlaProArgAsnIleArg 207
|||||
560 GCTGTACCAACACCTGGCATAGA...GCTGGCCCAAGAACATTAAGGT 606
|||||
207 lAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeu 224
|||||
607 GAACCTGCTTACCTGCTGCTTCAAGACTAGCTTGAAGGATGCTCT 656
|||||
224 rMetAspLysGluLysGluLysSerMetLysGluThrLeuArgIleArg 240
|||||
657 GGATGACACAGAAACAGAGAAACATGAAGAACCTGGGATAGA 706
|||||
241 ArgLeuGlyGluProGluAspCysAlaGlyIleValSerPheLeuCys 256
|||||
707 AGGTTAGGCGAG...CAGAGGATGCTGCTGCTGCTTCTTCTGCTGC 753
|||||
257 SerGluAspAlaSerIleIleThrGlyGluThrAlaValAlaGlyGly 273
|||||
754 TCTGAGATGCGATCACTACACTGGGGA...ACCTGTGTGTGGTGGGG 800
|||||
273 yThr 274
|||||
801 AAC 804

```

seq_name: gb_est2:BE741442

seq_documentation_block:
 LOCUS BE741442 755 bp mRNA linear EST 15-SEP-2000
 DEFINITION 601594426F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948247 5',
 mRNA sequence.
 ACCESSION BE741442
 VERSION BE741442.1 GI:10155434
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 755)
 AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM811 row: d column: 08
 High quality sequence stop: 752.
 Location/Qualifiers
 1. 755
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3948247"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average

FEATURES
 source

Insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using 2AP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 170 a 200 c 243 g 142 t
ORIGIN

alignment_scores:

Quality: 1169.00 Length: 234
Ratio: 4.996 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-866-034-2 x BE741442

Align seg 1/1 to: BE741442 from: 1 to: 755

```

7 LeuGlyLeuGlyAlaArgAlaTrpAsnSerValArgMetAlaSerSer1 23
3 CTAGGCTCTGTGCGGGGCTGTGAATTCGTGGGATGGCCAGCTCCGG 52
23 yMethrArgArgAspProLeuAlaAsnLysValAlaLeuValThrAla 40
53 GATGACCGCGCGGAGACCGCTCGCAATAAAGGTGGCCCTGTGTAACGGCCT 102
40 eThraspGlyIleGlyPheAlaIleAlaArgValLeuAlaGlnAspGly 56
103 CCACGCGAGGATCGGCTTCCGATCGCCCGCGCTTTGGCCAGAGCGGG 152
57 AlaHisValValSerSerArgLysGlnGlnAsnValAspGlnAlaVal 73
153 GCCCATGTGTCGTCCAGCAGCGGAGCAGATGTGACGACGCGCT 202
73 lAlaThrLeuGlnGlyLeuSerValThrGlyThrValCysHisV 90
203 GCCCAGCTGCGAGGGGAGGCTGAGCGCTGACGGGACCGCTGCCATG 252
90 aGlyLysAlaGlnAspArgGlnArgLeuValAlaThrAlaValLysLeu 106
253 TGGGAGAGCGGAGAGACCGGAGCGGCTGCGCCACCGCTGTGAAGCTT 302
107 HisGlyGlyIleAspIleLeuValSerAsnAlaValAsnProPhePh 123
303 CATGAGGATTCGATATCTCTAGTCTCAATGCTGCTCAACCCCTTCTT 352
123 eGlySerIleMetAspValThrGlnGlnValTrpAspLysThrLeuAsp 140
353 TGGAGGCTAATGATGATCTGAGGAGGTGGGAGCAAGACTCTGGACA 402
140 leAsnValLysAlaProAlaLeuMetThrLysAlaValAlaValProGln 156
403 TTATGTGTAAGGCCCGCCAGCCCTGATGACAAAGCGAGTGTGCCAATA 452
157 GluLysArgGlyGlySerValValIleValSerSerIleAlaAlaPh 173
453 GAGAAACGAGAGGAGGCTCAGTGTGATCGTCTTCATAGACGCCCTT 502
173 eSerProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeu 190
503 CAGTCCATCTCTGCTGCTCACTCTTACATGCTAGTAAACAGCCCTTGC 552
190 euGlyLeuThrLysThrLeuAlaIleGlnLeuAlaProArgAsnIleArg 206
553 TGGGCTTACCAAGACCTGGCCATAGAGCTGGCCCAAGAGACATTTAGG 602
207 ValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLe 223
603 CTGAACCTCCAGACCCGCGACTTATCAAGACTACCTTCACAGAGATCT 652
223 uTrpMetAspLysGlnLysGlnLysSerMetLysGlnThrLeuArgIleA 240
653 CTGATGTCACAAAGAAAAGAGAAAGCATGAAAGAACCTCGCGGATTA 702

```

240 rg 240
11
703 GA 704

seq_name: gb_est2:B1518701

seq_documentation_block:

LOCUS B1518701 967 bp mRNA linear EST 29-AUG-2001
DEFINITION 603062069p1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5211353 5',
mRNA sequence.
ACCESSION B1518701
VERSION B1518701.1 GI:15343493
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 967)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
Plate: LLM11530 row: 1 column: 18
High quality sequence stop: 841.

FEATURES

Location/Qualifiers

source

1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5211353"
/clone_lib="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH-MGC Library."

BASE COUNT 203 a 256 c 319 g 188 t 1 others
ORIGIN

alignment_scores:

Quality: 1148.00 Length: 288
Ratio: 4.284 Gaps: 13
Percent Similarity: 93.056 Percent Identity: 90.625

alignment_block:

US-09-866-034-2 x B1518701

Align seg 1/1 to: B1518701 from: 1 to: 967

```

1 MethIstLysAlaGlyLeuGlnGlyLeuGlyAlaArgAlaTrpAsnSerVal 17
27 ATGCACAAAGCGCGGCTGTGAGCTTGTGCGCGGCTTGAATTCGGT 76
17 lArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
77 CGCGATGGCGACGCTCCGGGATGACCGCGGAGACCCGCTCGCAATAAG 126
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50

```

```

127 TGCCCTGTACGGCTCCACGAGGAGTGGCTTGCATCGCCGG 176
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
177 CGTTGGCCAGAGGAGGCGCCATGTGCTGCTCAGCAGCCGGAAGCACA 226
67 nasValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
227 GAATGTGACAGGCGGTGGCCAGCTCCAGGGGAGGCGCTGAGCTGA 276
84 hrcGlyThrValCysHisValGlyLysAlaGlnAspArgLysVal 100
277 CGGCGACCGCTGCCATGTGGGGAAGCGGAGACCGGAGCGCTGGTG 326
101 AlathraValAlaLysLeuHisGlyLysAlaProAlaLeuMetThrLys 117
327 GCCACGCGCTGTAAGCTTCATGAGGTATCATATTCAGTCCCATGTC 376
117 AlaValAsnProPhePheGlySerIleMetAspValThrGluValT 134
377 TGCTGTACACCTTTCTTTGAGACATAATGATGTCTAGAGAGGTGT 426
134 rpaSPlyThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
427 GGGACGAGACTCTGACATTATGTGAAGGCCCAAGCCCTGATGACAAAG 476
151 AlaValAlaProGluMetGluLysArgGlyLysSerValAlaIleVal 167
477 GCACTGTGCGCCAGAAATGAGAAACGAGAGCGGCTCAGTGCATCGT 526
167 L.SerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsn 183
527 TGCTTCCATAGCAGCCTTCAGTCCATCTCCTGCTCAGTCCATTAACAAT 576
184 ValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluL 200
577 GTGAGTAACAGCCTTGCTGGGCTGACCAAGACCTGGCCATAGAGCT 626
200 euAlaProArg.AsnIleArgValAsnCysLeuAlaProGlyLeuIleLys 216
627 TGGCCCAAGGAGACATTAGGGTGAACCTGCTACCTGACTTATCAAA 676
216 sThrSerPheSerArgMetLeuTyr.MetAsp.Lys...GluLysGluI 231
677 GACTAGCTTACGACGAGATGCTGTGATTTGACCAAGGCCAACACAGCA 726
231 uSerMetLysGluThrLeuArgIleArg.ArgLeuGlyGluProGluAsp 247
727 CAGCATGAAGAAACCTCGGAGTAAACAGGTTAGCGCGACAGAGAT 776
248 CysAlaGly...IleValSerPheLeuCysSerGlu.AspAla.SerTyr 262
777 TGTGCTMGGCATCCGTGCTTTTCCCTGCTGCTGAAGATGCCAGCTTAC 826
263 IleThr.GlyGluThr.ValValValGly.GlyGlyThrPro 275
827 ATCACTGGGGAAACAGTGTGTGGTGGTGGAGGACCCCG 868
seq_name: gb_est2:BG696620
seq documentation block:
LOCUS BG696620 783 bp mRNA linear EST 07-MAY-2001
DEFINITION 602658918F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4801876 5',
VERSION BG696620
ACCESSION BG696620.1 GI:13961948
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 783)

```

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabs@email.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov
Plate: LLNL0695 row: d column: 05
High quality sequence stop: 781.
FEATURES
source
location/Qualifiers
1..783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4801876"
/lab_host="NCI_CGAP_Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 179 a 198 c 256 g 150 t
ORIGIN
alignment_scores:
Quality: 1147.00 Length: 262
Ratio: 4.552 Gaps: 3
Percent Similarity: 96.183 Percent Identity: 93.130
alignment block:
US-09-866-034-2 x BG696620
Align seq 1/1 to: BG696620 from: 1 to: 783
8 GlyLeuCysAlaArgAlaTyrPheSerValArgMetAlaSerSerGlyMet 24
3 GGCCTGTGCTGCTGGCAGCAAGTGGTGGGAGTGGCCAGCTCCAGAT 52
24 tThrArgArgAspProLeuAlaAsnLysValAlaLeuValThrAlaSerT 41
53 GACCCGCGGAGACCCGCTCAATATAGTGGCCCTGTAAAGGCTTCCA 102
41 hrasGlyIleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAla 57
103 CCGACGGGATCGGCTTGGCCATGCCCCGGCGTTGGCCAGACAGGGCC 152
58 HisValValSerSerArgLysGlnGlnAsnValAspGlnAlaValAl 74
153 CAGCTGTGCTGACACCGCGAAGCAGACAGAAATGTGACCAAGCGGTGC 202
74 aThrLeuGlnGlyGlyLeuSerValThrGlyThrValCysHisValG 91
203 CAGCTGACAGGGGAGAGGCTGACGCTGAGCGCACCTGTGCCATGTGG 252
91 LysAlaGlnAspArgGluArgLeuValAlaThrAlaValLysLeuHis 107
253 GGAAGGCGGAGACCGCGAGCGCTGTGTCATAGGCTGTGAAGCTTCAT 302
108 GlyGlyLysPheLeuValSerAsnAlaValAlaAsnProPheGly 124
303 GGAAGTATCATATCTAGTCCAAATGCGTCAACCTTTCTTTGG 352
124 YserIleMetAspValThrGluGluValTrrpaSPlyThrLeuAspIle 141
353 AACCTTAATGATGTCAACGAGAGGTGTGGCAAGACTCTGACACTTA 402
141 snValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetGlu 157

```

403 ATGTGAGGCCCGCCGCTGATGACAAAGGCAAGTGTGCCAGAAATGAG 452

158 LysArgGlyGlyIleValIleValIleValSerSerIleAlaIleAlaPhe 174
 453 AAMCGAGGAGGGGGCTGCTGATGATGCTGCTTCATAGCAGCTTCAG 502

174 rProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeuLeu 191
 503 TCCATCTCCCTGGCTTCAGTCCCTTACAAATGTCAATAAAGACCTTCCTG 552

191 LyeuThrLysThrLeuAlaIleGluLeuAlaProArgAsnIleArgVal 207
 553 GCCTCAACAATACCTGGCCATAGAGCTGCCCCAAGAAACATTAGGGT 602

208 AsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeu 224
 603 AACTGCTTGCACCTGACTTATCAGACTAGCTTACAGAGATGCTG 651

224 p.MetAspLysGluLys.GluGluSerMetLysGluThrLeuArgIle 240
 652 GATTGCAAGAGGAAACAGAGCAACATGAAAGAAAGCCCTGCGATAG 701

240 ArgLeuGlyGluProGlyLysPheCysAlaGlyIle.ValSerPheLeu 256
 702 AAGGTTAGCGGAGCCAGAGATTGTCTGCGATGCTCTTCCTGCTG 751

257 SerGluAspAlaSerTyrIleThrGlyGlu 266
 752 TCTGAGATGCCAGTACATCACTGGGAA 781

seq_name: gb_est2.B1335522

seq_documentation_block:
 LOCUS B1335522 724 bp mRNA linear EST 30-JUL-2001
 DEFINITION 602997109P1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5139338 5',
 mRNA sequence.
 ACCESSION B1335522
 VERSION B1335522.1 GI:15020179
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 724)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/MLN at:
 http://image.llnl.gov
 Plate: LRAM1343 row: a column: 03
 High quality sequence start: 11
 High quality sequence stop: 722.
 Location/Qualifiers
 source
 1. 724
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5139338"
 /clone_lib="NIH_MGC_12"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 151 a 196 c 239 g 138 t
 ORIGIN

alignment_scores:
 Quality: 1105.00 Length: 230
 Ratio: 4.868 Gaps: 2
 Percent Similarity: 98.696 Percent Identity: 98.261

alignment_block:
 us-09-866-034-2 x B1335522

Align seg 1/1 to: B1335522 from: 1 to: 724

1 MethiLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVal 17
 37 ATGCACAGAGCGGGGCTGATGCTCTGTCGCGGGCTTGGAATTCGCT 86
 17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
 87 GCGGATGCGCCAGCTCGGGATGACCCCGGACCCGCTGCAAAATAGG 136
 34 AlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 137 TGGCCCTGGTACGGCTCCACCGAGGATCGGCTTCGCCATCGCCGG 186
 51 ArgLeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGlnI 67
 187 CGTTTGGCCAGAGCGGGCCCATGTGTCGTCAGACGCGGAGAGAGCA 236
 67 AsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValTr 84
 237 GAATGTGACACAGCGGTGCGCCAGCTGCGAGCGGAGGGCTGAGCGTGA 286
 84 hArgLysThrValCysHisValGlyLysAlaGlnAspArgGluArgLeuVal 100
 287 CGGGACCGCTGCTCCATGTGGGAGAGCGGAGACCGGAGCGGCTG 336
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
 337 GCCACGCGCTGAAAGCTTCATGAGAGTATCATCTTACATCCCAATGC 386
 117 AlaValAsnProPhePheGlySerIleMetAspValThrGluGluVal 134
 387 TGTGTTCAACCCCTTCTTGGAGCATTAATGATGTCACAGAGAGGTGT 436
 134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 437 GGGACAGACTGTGACATTAAATGTGAAGGCCCGCAGCTGATGACAAAG 486
 151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleVal 167
 487 GCAGTGTGTCAGCAAAATGAGAAACAGAGAGCGGCTCACTGATGCTGT 536
 167 LSerSerIleAlaIleAlaPheSerProSerProGlyPheSerProTyrAsn 184
 537 GTCCTTCATAGCAGCTTCAGTCCATCTCTGCTTACAGCTTACATG 586
 184 aLysLysThrAlaLeuLeuGlyIleuThrLysThrLeuAlaIleGluLeu 200
 587 TCAGTAAACAGCCCTTCTGCGCTGACCAAGACCCGCGGCAATAGAGCTG 636
 201 AlaProArgAsn.IleArgVal.AsnCysLeuAlaProGlyLeuIleLys 216
 637 GCCCAAGAGAACCATTTAGGGTTGACCTGCTAGACCTTGACATTATCAAG 686
 217 ThrSerPheSerArgMetLeuTrpMetAspLysGlu 228
 687 ACTAGCTTCAGCAGAGATGCTGTGATGGACAAGGAA 722

seq_name: gb_est2.BG399750

seq_documentation_block:
 LOCUS BG399750 817 bp mRNA linear EST 12-MAR-2001
 DEFINITION 602441351F1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4556940 5',

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10386 row: e column: 16
 High quality sequence stop: 777.
 Location/Qualifiers

FEATURES

1..956

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4508151"
 /clone_id="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: PCMV-SPORT6; Site: 1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 212 a 251 c 324 g 167 t 2 others
 ORIGIN

alignment_scores:

Quality: 1102.50 Length: 254
 Ratio: 4.446 Gaps: 2
 Percent Similarity: 97.638 Percent Identity: 94.882

alignment_block:

us-09-866-034-2 x BG257840

Align seg 1/1 to: BG257840 from: 1 to: 956

```

1 Methislysalaglyleuenglyleucysalaargalaatrpaserva 17
|||||
7 ATGCACAAAGCGGGCTGTAGCCCTCTGTGCCCGGCTTGAATTCGCT 56
|||||
17 lartmetaserseglmetlthararqaspproleualasnlvsv 34
|||||
57 :CGGATGGCCAGCTCCGGGATGACCCCGCGGACCGCTCGCAATTAAG 106
|||||
34 alalaleuvalthralaserthrpsglyleglyphealalealarg 50
|||||
107 TGGCCCTGTAGCGGCTCCACCGAGCGGATCGGCTTCGCATCGCCGG 156
|||||
51 Argylealalaglnaspolyahisvalvalaserarplysngl 67
|||||
157 GATTGGCCAGGAGCGGCCCATGTGTCAGACGCGGAGAGAGAGCA 206
|||||
67 nasnvalaspglnalavalathrlenglnglyuglyleuservalt 84
|||||
207 GAATGTGGACCAAGCGGTGGCCAGCTGAGGGGGGGGCTGAGCGTGA 256
|||||
84 hnglythralcysnhsvalglylsalagluaspargluarqleuval 100
|||||
257 CGGGACCGCTGTGCCATGTGGGGAAGCGGAGACCGGGAGCGGCTGTG 306
|||||
101 Alathralavalysleuhsclglylleasplleleuvalserasna 117
|||||
307 GCCAGCGCTGTAGAGCTCATGTATCGATATCTAGTCTCCANGC 356
|||||
112 alalavalasnphepheglyserllemetaspyaltrpgluglualt 134
|||||
357 TCGTGTCAACCCCTTCTTGGAGCATATGTGATGTCACATGAGAGCTGT 406
|||||
134 rPasplysthreusaplleasvalysalaproalaleumetthrvs 150
|||||
407 GGGACAGAGCTCTGGACATTATGTGAAGGCCCGACCTGTATGACAAAG 456
|||||
151 Alavalvalproglumetglulysargglyglyservalvalileva 167
|||||
457 GCAGTGTGCCAGAAATGAGAGAGAGGCGGCTCAGTGTATCTGT 506
|||||
167 lserlrelealalapheserproserproglypheserprotyrasnv 184
|||||

```

```

507 GTCTTCATAGCAGCCTTCAGTCCATCTCTGGCTTCAAGCTTACAAAG 556
184 alserlysthralaleuenglyleuthrlysthreualallegluen 200
|||||
557 TCAGTAAACAGCCCTTGTCT. GGGCTGACCAAGACCTTGGC. ATAGAGCTG 604
|||||
201 Alaproargsnlleargvalasncysleualaproglyleuilelysth 217
|||||
605 GCCCAAGAGAACCTTAGGTAAGTCCCTTAGACCTGTATCAAGAC 654
|||||
217 rserpheserarmetleutrpmetasplysglyuglusermetl 234
|||||
655 TGGCTTCAGAGAGATGCTCCGGATGACAGAGAACAGAGAGAGC...T 701
|||||
234 ysglu.Threuarqgleaargargleuglygluprogluaspcysalagl 250
|||||
702 GAGAGAAACCTGCGGATGAAGAANGTTAGGAGACCGAG. GATTGTCTGTG 750
|||||
250 yllevalser 253
|||||
751 CATCGTGTCTT 760

```

seq_name: gb_est2:BG436827

seq_documentation_block:

LOCUS BG436827 695 bp mRNA linear EST 14-MAR-2001
 DEFINITION 60248812P1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621082 5',
 mRNA sequence.

ACCESSION BG436827
 VERSION BG436827
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1- (bases 1 to 695)
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bhs-remail.nih.gov
 Tissue Procurement: DCRD/DMP/Gazdar

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: L10C1193 row: g column: 03
 High quality sequence stop: 695.

FEATURES

1..695

source
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4621082"
 /clone_id="NIH_MGC_18"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pORB7; Site: 1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."
 BASE COUNT 144 a 183 c 231 g 137 t
 ORIGIN

alignment_scores:

Quality: 1097.00 Length: 225
 Ratio: 4.876 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.556

alignment block:
US-09-866-034-2 x BG436827

Align seg 1/1 to: BG436827 from: 1 to: 695

```

1 Methislyslaglyleuenglyleucysalargalatrpsnserva 17
2 ATGCACAAAGGGGGGCTGCTAGGCTCTGTCGCCGGGCTTCGAATTCGCT 71
17 largmetalaasersearlymetthrargaspptoleualasnlvsy 34
72 GCGGATGGCCAGCTCCGGGATGAGACCCCGGAGACCCCTGCCAAATAGG 121
34 alalaleuvalthralserthraspolylllelyphealaleaary 50
122 TGGGCTTGTTAGACGGCTCCACCGAGGGATGCGCTTCGCCATGCGCCGG 171
51 Argleualaglnaspglalahisvalvalsersearlyglncl 67
172 CGTTTGGCCCAAGAGGGGGCCCATGTGCTCAGCAGCGGAAAGCAGCA 221
67 nasnvalaspglalavalathrleuglnlyglnlyleusenervalt 84
222 GAATGTGACACAGCGGCTGCGCCACGCTCCAGGGGAGGGGCTGACCTGA 271
84 hrelgylthrvalcyslhisvalglylsalaglnasparglunargleuval 100
272 CGGGCACCCGTGTGCATGTGGGAGAGCGGACCGGAGCGGCTGTG 321
101 Alathralavalalysleuhsiglylilasplileuvalserasna1 117
322 GCCACGGCTGTGAAGCTTCATGAGGTATGATATCTAGTCTCAATGC 371
117 aalavalasnpheptheglyserilemetaspvalthrsluvalt 134
372 TGCTGTCAACCTTCTTGGAAACATATGATGTCTCAGAGAGGTGT 421
134 rpaasplysthrleuasplileasnvallysalaalaleuvalmetthlys 150
422 GGGACAAAGCTCTGACATTAATGTGAAGGCCCGGACCTGCATGACAAAG 471
151 Alavalvalproglumetglulysargglylygylservalallea 167
472 GCGTGTGTGCCAGAAATGAGAAACGAGGAGCGGCTCAGTGTGATCTGT 521
167 lserseerlealalapheserproserproglypheserprotyrasny 184
522 GTCTTCATAGACGCTTCAGTTCATCTCTGCTTCAGCTTCATCAATG 571
184 alserlysthralaleuenglyleuthrlysthrleualaleglnleu 200
572 TCAGTAAACAGCCTGTGGGCTCAGCAAGACCTGTGCTATGAGCTG 621
201 Alaproarasnlllearyvalasnycysleualaproglyleu1leysth 217
622 GC.CCAAGAACATTAGGCTGAACTCCTAGACACTGACTTATCAAGAC 670
217 rserpheserargmetleutrimet 225
671 TAGCTTACACAGAGATGCTGTGATG 695

```

seq_name: gb_est2.BE563882

seq_documentation_block: 975 bp mRNA linear EST 15-AUG-2000

LOCUS BE563882 601335069P1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689176 5',

DEFINITION mRNA sequence.

ACCESSION BE563882

VERSION BE563882.1 GI:9807602

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 975)

AUTHORS

NIH-MGC <http://mgi.mgi.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNU at: <http://image.llnl.gov>

FEATURES

source

plate: LLCM383 row: 1 column: 17
High quality sequence start: 82
High quality sequence stop: 734.
Location/Qualifiers
1..975

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3689176"

/clone_lib="NIH_MGC_39"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7. Site.1: XhoI; Site.2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

199 a 254 c 319 g 203 t

alignment_scores:

Quality: 1096.00 Length: 266
Ratio: 4.455 Gaps: 6
Percent Similarity: 92.481 Percent Identity: 88.722

alignment block:

US-09-866-034-2 x BE563882

Align seg 1/1 to: BE563882 from: 1 to: 975

```

1 Methislyslaglyleuenglyleucysalargalatrpsnserva 17
2 ATGCACAAAGGGGGGCTGCTAGGCTCTGTCGCCGGGCTTCGAATTCGCT 112
17 largmetalaasersearlymetthrargaspptoleualasnlvsy 34
113 GCGGATGGCCAGCTCCGGGATGAGACCCCGGAGACCCCTGCCAAATAGG 162
34 alalaleuvalthralserthraspolylllelyphealaleaary 50
163 TGGGCTTGTTAGACGGCTCCACCGAGGGATGCGCTTCGCCATGCGCCGG 212
51 Argleualaglnaspglalahisvalvalsersearlyglncl 67
213 CGTTTGGCCCAAGAGGGGGCCCATGTGCTCAGCAGCGGAAAGCAGCA 262
67 nasnvalaspglalavalathrleuglnlyglnlyleusenervalt 83
263 GAATGTGACACAGCGGCTGCGCCACGCTCCAGGGGAGGGGCTGACCGTG 312
84 ThrGlyThrValCysHisValGlyLysAlaGlnAspArgLunArgLeuVa 100
313 ACGGGCACCGTGTGCATGTGGGAGGCGGAGAGACCGGACGGGCTGTGT 362

```

```

100 1A1ATrAlaValLysLeuHISGLYGLYLeAspIleLeuValSerAsnA 117
|||||
363 GGCACAGGCGTGTGAAGCTTCATGAGATATCATATCCATGCTCCAAATG 412
|||||
117 1A1AlaValAsnProPhePheGlySerIleMetAspValThrGluGluVal 133
|||||
413 CTGCTCTCAACCTTCTTGTGAAGCATATATGATGTCACCTGAGAGAGTG 462
|||||
134 TrpAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
|||||
463 TGGGACAAAGACTGTGCATTAATGTGAAGGCCCCAGCCCTGATGACAAA 512
|||||
150 sAlaValAlaProGlu.MetGluLysArgGlyGlySerValIle 166
|||||
513 GGCAGTGTGCCAGAAATGTGAGAAAGAGAGCGGCTCAGTGTATC 562
|||||
167 ValSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAs 183
|||||
563 GTGCTCTTCATAGCAGCTTCATCATCTCCTGCTTCAGTCTTACAA 612
|||||
183 nValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 200
|||||
613 TGTCACTAAACAGCCTTGCTGGGCTGACCAAGACCCCTGGCATAGAG 661
|||||
200 euAla.ProArgAsnIleArg.ValAsnCysLeuAlaProGlyLeuIle 216
|||||
662 ..CTGGCCCGAGAACATTAAGGGTGAAGCTGACACCGAGCTTATCA 709
|||||
216 yThrSerPheSerArg...MetLeuTrpMetLysPheGlyGluGlu 231
|||||
710 ACAGTACGCTTCAGACGAGGATGCTGTGATTGTGACGCGAAACAGAGAA 759
|||||
232 SerMetLysGluThrLeuArgIleArgArgLeuGlyGluProGluSpc 248
|||||
760 ACCCTCAAGAACCCGCGGATTAAGAAAGCTTAGCGGACCGCAGAGATTG 809
|||||
248 sAlaGlyIleValSerPheLeuGlySerGluAspAlaSer 261
|||||
810 TCCGTGCAATCGTGTCTTCTGTGCTGCAAGATCCAGCT 849

```

seq_name: gb_est2:B1160793

seq_documentation_block:

```

LOCUS      B1160793              806 bp      mRNA      linear      EST 05-JUL-2001
DEFINITION 602864883F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5018877 5',
            mRNA sequence.
ACCESSION  B1160793
VERSION    B1160793.1  GI:14620794
KEYWORDS   EST.
SOURCE     human.

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 806)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: rgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LCM1833 row: e column: 22
            High quality sequence stop: 708.

```

FEATURES

```

source
1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5018877"

```

```

/clone.lib="NIH_MGC_42"
/tissue.type="epithelioid carcinoma cell line"
/lab.host="DH108 (phage-resistant)"
/note="Organ: pancreas; Vector: pOMB7; site: 1: XhoI;
site: 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. 1"

```

```

BASE COUNT      182 a      203 c      267 g      154 t
ORIGIN

```

alignment_scores:

```

Quality: 1076.00      Length: 272
Ratio: 4.321          Gaps: 8
Percent Similarity: 91.544      Percent Identity: 88.235

```

alignment_block:

US-09-866-034-2 x B1160793

Align seg 1/1 to: B1160793 from: 1 to: 806

```

1 MetHisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
|||||
5 ATGCACAAAGCGGGGCTGCTAGGCGCTTGTGCCGCGCTTGGAATTGCGT 54
|||||
17 LArgMetAlaSerSerGlyMethThrArgArgAspProLeuAlaAsnLys 34
|||||
55 GCGGATGCGCAGCTCGGGATGACCCGCGGACCGCTGCAAAATGAG 104
|||||
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
|||||
105 TGGCCCTGCTTAACGGCGCTCCACCGAGGAGTCCGCTTGGCATCGCCGG 154
|||||
51 ArgLeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGln 67
|||||
155 CGTTTGGCCGAGGAGCGGGCCCATGTGCTGCAGACGCGGAAAGCAGCA 204
|||||
67 nasValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
|||||
205 GAATGTGGACAGCGGCTGGCCAGCTGCAGGGGGAGGGCTAGGCTGA 254
|||||
84 hArgLysThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
|||||
255 CGGGCAACCGTGTCCATGTGGGAGAGCGGAGACCGGAGCGGCTGCTG 304
|||||
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnA 117
|||||
305 GCCACGCGCTGTGAAGCTTCATGAGGATATCATCTCACTGCTCAATGC 354
|||||
117 aAlaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
|||||
355 TGTCTGTCAACCTTCTTGTGAAGCATATATGATGTCACCTGAGAGAGTG 404
|||||
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
|||||
405 GGCACAAAGACTGTGCATTAATGTGAAGGCCCAAGACTGATGACAAAAG 454
|||||
151 AlaValAlaProGluMetGluLysArgGlyGlySerValValIleVal 167
|||||
455 GCAGTGTGCCAGAAATGTGAAGAGAGAGCGGCTCAGTGTATCGT 504
|||||
167 lSerSer.IleAla.AlaPheSerProSerProGlyPheSerProTyrAs 183
|||||
505 GTCTTCCATATAGCCAGCTTCATCATCTCCTGCTTCAGTCTTACAA 554
|||||
183 nValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 200
|||||
555 TGTCACTAAACAGCCTTGCTGGGCTGACCAAGACCCCTGGCCATAGAGC 604

```


OM of: US-09-866-034-2 to: GenEmbl.* out_format : pfs

Date: May 9, 2002 3:07 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-frame.p2n.model -DEV-xlp
-O/-cgn2_l/usprto.spool/us09866034/runat_08052002_081343_16433/app_query.fasta_1.338
-DB-GenEmbl -OPMT-fastap -SUFFIX=2n.rge -GAPOP=12.000
-GAEXP=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.000
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-blossum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR-US09866034.ecgnl.1.6980 -NCPU=6 -ICPU=3 -LONLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLDPX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-866-034-2
Query length: 278
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1898.670000
```

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr:BC003019	+ 1399.00	1953.91	1.8e-100	1258	BC003019 Homo sapiens, peroxis
gb_pat:AX201380	+ 1399.00	1953.71	1.8e-100	1283	AX201380 Sequence 59 from Pat
gb_pr:AE0045131	+ 1399.00	1953.39	1.9e-100	1325	AB045131 Homo sapiens humnDR
gb_pr:AF001870	+ 1395.00	1948.31	3.6e-100	1258	AK001870 Homo sapiens cDNA FL
gb_pr:AF004127	+ 1391.00	1942.52	7.7e-100	1281	AF004127 Homo sapiens peroxis
gb_com:AB062157	+ 1154.00	1614.10	1.5e-81	880	AB062157 Sus scrofa mRNA for c
gb_to:AB045132	+ 1130.50	1577.15	1.7e-79	1312	AB045132 Mus musculus mounDR
gb_com:AB045133	+ 1123.00	1567.44	6.0e-79	1212	AB045133 Oryctolagus cunicul
gb_to:BC003484	+ 1117.00	1558.85	1.8e-78	1234	BC003484 Mus musculus, Simla
gb_to:AB062155	+ 1073.00	1501.45	2.8e-75	806	AB062155 Rattus norvegicus, Simla
gb_pr:AY071856	+ 923.00	1289.01	1.9e-63	1003	AY071856 Homo sapiens, NMDP-de
gb_pr:BC007339	+ 833.50	1161.55	2.4e-56	1226	BC007339 Homo sapiens, short-c
gb_pr:HS031875	+ 824.50	1147.31	1.5e-55	1442	U31875 Human Hep2 protein mRN
gb_in:AY069779	+ 664.00	924.95	3.9e-43	1113	AY069779 Drosophila melanogast
gb_in:AC015075	+ 619.00	827.53	9.8e-38	34084	AC015075 Drosophila melanogast
gb_in:AC009801	+ 619.00	816.25	4.1e-37	104808	AC009801 Drosophila melanogast
gb_in:AY035106	+ 588.50	819.36	2.8e-37	178360	AY035106 Arabidopsis thaliana
gb_in:CGR01A11	+ 546.00	723.90	5.8e-32	38822	AC084173 Caenorhabditis brig
gb_in:AE007111	+ 462.50	613.66	8.0e-26	19741	AE007111 Mycobacterium tuberc
gb_ba:MTV002	+ 462.50	603.12	3.1e-25	56414	AI008967 Mycobacterium tuberc
gb_ba:AE005197	+ 454.00	608.54	1.5e-25	10029	AE005197 Caulobacter crescent
gb_ba:AE004169	+ 413.50	550.63	2.6e-22	11241	AE004169 Streptomyces aerugin
gb_pat:655145	+ 409.00	542.96	6.9e-22	12879	AE004609 Pseudomonas aerugin
gb_ba:AE003189	+ 404.00	564.53	4.4e-23	747	E65145 Process for producing of
gb_ba:BSU05433	+ 400.00	500.21	1.7e-19	298050	AP003189 Clostridium perfrin
gb_ba:BSUB0009	+ 399.00	500.96	1.5e-22	2678	U59433 Bacillus subtilis, plax
gb_ba:AE001811	+ 396.50	527.77	4.9e-21	10206	E29112 Bacillus subtilis, com
gb_in:CEP54F3	+ 396.00	522.45	9.6e-21	16173	AE001811 Thermococcus maritima
gb_ba:AL646077	+ 392.00	492.16	9.6e-21	189050	E279696 Caenorhabditis elegans
gb_ba:SC2G61	+ 391.50	506.55	7.4e-20	42071	AL359949 Streptomyces coelic
gb_ba:AL646062	+ 389.50	488.45	7.5e-19	193050	AI646062 Ralstonia solanace
gb_ba:AT321122	+ 388.00	516.21	2.1e-20	9858	AF321122 Streptomyces antibiot
gb_ba:AT324838	+ 388.00	503.52	1.3e-19	34869	AF324838 Streptomyces antibiot
gb_ba:AT322256	+ 388.00	502.29	1.1e-19	39428	AF322256 Streptomyces antibiot
gb_ba:AE0001515	+ 385.00	481.81	1.8e-18	303249	AP001515 Bacillus halodurans
gb_ba:AE000752	+ 385.00	509.70	4.9e-20	12395	AE000752 Aquifex aeolicus, seq
gb_ba:AE003001	+ 385.00	476.29	3.6e-18	345783	AP003001 Mesorhizobium loti
gb_ba:AP003009	+ 383.50	474.37	4.6e-18	339681	AP003009 Mesorhizobium loti

```
seq_name: gb_pr:BC003019
seq_documentation_block:
LOCUS BC003019 1258 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, peroxisomal short-chain alcohol dehydrogenase, clone
MGC:4052 IMAGE:2822884, mRNA, complete cds.
ACCESSION BC003019
VERSION BC003019.1 GI:12804320
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1258)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
```

REMARK
CONTACT: MGC help desk
http://mgc.ncl.nih.gov

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Email: cgabs-rt@mail.nih.gov
Tissue Procurement: DCTD/DP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chin, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Seedei, Jacqueline
Schlein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuydamduy, Marco Mairra.

FEATURES
source
1. 1258
/organism="Homo sapiens"
/db_xref="locusID:10501"
/db_xref="taxon:9606"
/clone="MGC:4052 IMAGE:2822884"
/tissue_type="Liver, small cell carcinoma"
/clone_id="NIH MGC.7"
/lab_host="DH10B-R"
/note="vector: pOTB7"
3. 839
/codon_start=1
/product="peroxisomal short-chain alcohol dehydrogenase"
/protein_id="AAH03019.1"
/db_xref="GI:12804321"
/translation="MKAAGLGLCARANNSYRMASGTRDPLANKVALVYASTDGI
GVALARRADAGAAHVAVSSRRQAVNDVAVTLOEGSVTCVHYKAEDEKRLVAT
AVKLHGIDILVSNAAVAVPFGSDIMDYEEVMDLTDINVAAPALMKAVPDEKRG
GGSVIVASSIAFSPSPGFSYNSKTALELTFTLAIETLAPRINVCNAPGLIKTS
FSRMLMDKKEEEMKELTARRRGEPEBDCAIGYFLCSDEASITTEGTYYVGGTFS
RL"

CDS

BASE COUNT 296 a 318 c 390 g 254 t
ORIGIN

alignment_scores:

Quality: 1399.00 Length: 278
 Ratio: 5.032 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-866-034-2 x BC003019

Align seg 1/1 to: BC003019 from: 1 to: 1258

```

1 MethistylsalaglyleuLeuGlyLeuGlyCysAlaArgAlaTrpAsnSerVa 17
  |||||||
3 ATGCACAAGGGGGGCTCTAGGCTCTGCTGCTCCGGGCTTGGAATTCGCT 52
  |||||||
17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnIysV 34
  |||||||
53 GCGGATGGCCAGCTCCGGGATGACCCCGCGGAGCCCTCCCAATAAAG 102
  |||||||
34 alAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
  |||||||
103 TGGCCCTGGTAAAGGCTCCACCAAGGATGCGCTTGCCATGCGCCGG 152
  |||||||
51 ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgIysGlnI 67
  |||||||
153 CGTTTGGCCAGAGAGGGGCCATGTGTCTCAAGACCGGAGACAGCA 202
  |||||||
67 nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
  |||||||
203 GAATGTGACCAAGGCGGTGCGCCAGCTCCAGGGGAGGGCTGAGCGTGA 252
  |||||||
84 hrGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
  |||||||
253 CGGGCACCCTGTGCCATGTGGGGGAAGCGGAGACCGGGAGCGCTGTG 302
  |||||||
101 AlaThrAlaValLysLeuHisGlyLysIleAspIleLeuValSerAsnAl 117
  |||||||
303 GCCACGGCTGTGAAGCTTCATGAGAGGTATCGATATCTGCTCAATGC 352
  |||||||
117 aaIaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
  |||||||
353 TGGCTTCACCCCTTCTTGGAGCATATATGATGTCTCATGAGAGGCT 402
  |||||||
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
  |||||||
403 GGGACAAGACTCTGACATTAATGTGAAGGCCCAAGCCCTATACAAAG 452
  |||||||
151 AlaValValProGluMetGluLysArgGlyLysIleSerValValIleVa 167
  |||||||
453 GCAGTGTGGCCAGAAATGAGAAAGAGAGGCGGCTCAGTGTGATGCT 502
  |||||||
167 lSerSerIleAlaIleAlaPheSerProSerProGlyPheSerProTyrAsnV 184
  |||||||
503 GTCTTCCTAGACGCTTCAGTCCATCTCTGCTTCAGTCTTCAATAG 552
  |||||||
184 alSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
  |||||||
553 TCAGTAACAAGCCTGTGCTGGCTGACCAAGACCTGCGCATAGAGCTG 602
  |||||||
201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysTh 217
  |||||||
603 GCCCCAAGACATTAAGGTGAAGTGTGCTTACCTGAGCTTATCAAGAC 652
  |||||||
217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLeuSerMetL 234
  |||||||
653 TAGCTTCAGCAGATGCTCTGTGATGACAAAGAAAGAAAGAAAGCAATCA 702
  |||||||
234 ySGluThrLeuArgIleArgArgLeuGlyLysProGluAspCysAlaGly 250
  |||||||
703 AAGAAACCTCGGATAGAGGTTAGCGAGCCAGAGAGATTGCTGCGC 752
  |||||||
251 lIleValSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyLysTh 267
  |||||||

```

753 ATCGTCTTTCCTGCTCTGAGATGCCAGCTACATCATCTGGGGAAC 802
 267 rValValValGlyGlyGlyThrProSerArgLeu 278
 803 AGTGTGTGCTGTGGAGGAGACCCCTCCGCTC 836

seq_name: gp_Pat:AX201380

seq_documentation_block:

LOCUS AX201380 1283 bp DNA linear PAT 30-AUG-2001
 DEFINITION Sequence 59 from Patent WO0153486.
 ACCESSION AX201380
 VERSION AX201380.1 GI:15391185
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 1283)
 Ashkenazi, A.J., Goddard, A., Godowski, P.J., Gurney, A.L.,
 Hillan, R.J., Marsters, S.A., Pan, J., Pittl, R.M., Roy, M.A., Smith, V.,
 Stone, D.M., Watanabe, C.K. and Wood, W.I.
 Compositions and methods for the treatment of tumour
 Patent: WO 0153486A 59 26-JUL-2001;
 Genentech, Inc. (US)

FEATURES
 Location/Qualifiers

source
 1..1283
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 292 a 327 c 401 g 263 t
 ORIGIN

alignment_scores:
 Quality: 1399.00 Length: 278
 Ratio: 5.032 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-866-034-2 x AX201380

Align seg 1/1 to: AX201380 from: 1 to: 1283

```

1 MethistylsalaglyleuLeuGlyLeuGlyCysAlaArgAlaTrpAsnSerVa 17
  |||||||
36 ATGCACAAGGGGCTCTAGGCTCTGCTGCTCCGGGCTTGGAATTCGCT 85
  |||||||
17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnIysV 34
  |||||||
86 GCGGATGGCCAGCTCCGGGATGACCCCGGGAGCCCTGCCAATAAAG 135
  |||||||
34 alAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
  |||||||
136 TGGCCCTGGTAAAGGCTCCACCAAGGATGCGCTTCCATGCGCCCGG 185
  |||||||
51 ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgIysGlnI 67
  |||||||
186 CGTTTGGCCAGAGAGGGGCCATGTGTGCTCAAGACCGGAGACAGCA 235
  |||||||
67 nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
  |||||||
236 GAATGTGACCAAGCGCGGTGCGCCAGCTCCAGGGGAGGGCTGAGCGTGA 285
  |||||||
84 hrGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
  |||||||
286 CGGGCACCCTGTGCCATGTGGGGGAAGCGGAGACCGGAGCGCTGTG 335
  |||||||
101 AlaThrAlaValLysLeuHisGlyLysIleAspIleLeuValSerAsnAl 117
  |||||||
336 GCCACGGCTGTGAAGCTTCATGAGAGGTATCGATATCTGCTCAATGC 385
  |||||||
117 aaIaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
  |||||||

```

```

386 TGCTGTCAACCCCTTCTTGGAGCATTAATGATGTCCACTGAGAGGTGT 435
134 rpsplysthleuasplleasvalysalaproalaueuethrlys 150
436 GGGACAGAGCTGTGACATTAAATGAGAGCCGCCAGCTGATGACAAAG 485
151 AlalaValaProglumetglulysarglyglySerValaIleVa 167
486 GCAGTGTGTCAGAAATGAGAAACAGAGAGCGGCTCAGTGTGATCGT 535
167 lSerSerIleAlaIleAlaPheSerProSerProgluPheSerProtyrAsn 184
536 GTCTTCATAGCAGCCTTCATGCTCATCTCTGCTGCTCAGTCTTACAAG 585
184 alSerlysthAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
586 TCAGTAAACAGCCTTGTGGGCTGACCAAGACCTGCGCATAGAGCTG 635
201 AlaproArgAsnIleArgValaAsnCysLeuAlaProgluLeuIleLysTh 217
636 GCCCAAGAGACATTAGGTGAGTGAACCTGACCACTGGACTTATCAGAG 685
217 rSerPheSerArgMetLeuThrMetAspLysglulysgluIleSerMet 234
686 TAGCTTCACAGAGATGCTCTGATGAGCAAGAAAGAGAAAGCATGA 735
234 ysgLThrLeuArgIleArgArgLeuGlyGluProgluAspCysAlaGly 250
736 AAGAAACCTGCGGATTAAGAGATTAGCGACGACGAGATTGCTGTGCG 785
251 lIleValSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluTh 267
786 ATCGTGTCTTCTGCTGCTGCTGAACATGCCACTCATCTACTGGGAAAC 835
267 rValaValaIleGlyGlyThrProSerArgLeu 278
836 ACTGTGTGTGGTGAAGAACCCGCTCCGCTC 869
seq_name: gb_AB045131

```

```

seq_documentation_block:
LOCUS AB045131 1325 bp mRNA linear PRI 05-DEC-2000
DEFINITION Homo sapiens humNRDR mRNA for NADPH-dependent retinol
dehydrogenase/reductase, complete cds.
ACCESSION AB045131
VERSION AB045131.1 GI:11559411
KEYWORDS
SOURCE Homo sapiens liver cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS Furukawa,A., Ohnishi,T., Huang,D., Araki,N. and Ichikawa,Y.
TITLE cDNA cloning and characterization of peroxisomal short-chain
dehydrogenase / reductase that reduce all-trans retinal to retinol
JOURNAL Unpublished
2 (bases 1 to 1325)
REFERENCE
AUTHORS Furukawa,A., Huang,D., Ohnishi,T. and Ichikawa,Y.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2000) Aizo Furukawa, Kagawa Medical University,
Faculty of Medicine, Department of Biochemistry: Miki-cho, Ikenobe
1750-1, Kita-gun, Kagawa 761-0793, Japan
( E-mail: afurukawa@kms.ac.jp, Tel:81-87-891-2104, Fax:81-87-891-2105)
FEATURES
SOURCE
1. 1325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
134..916
/gene="humNRDR"
134..916
/gene="humNRDR"
/codon_start=1

```

```

gene
CDS
/codon_start=1

```

```

/product="NADPH-dependent retinol dehydrogenase/reductase"
/protein_id="BAB18775.1"
/db_xref="GI:11559412"
/translacion="MASSGMPREDPLANKVALYTASTDGIGFATRRRLAODCAHYVS
SRKQONVDAVATLOGECISVTGYCHGKEDRRLVATKHLGGIDILVSNAYN
PFGSIMYTEREWDKTDINKAPALMTKAVPMERGGGSVVIVSIAFSPG
FSPYVSVTALGLTKRLAIELAPRINIVNCLAPOLIKTISFSRMLMDKREESMET
LRIIRLGEPEDCAGIVSFLCSRDASYITGETVYVVGCTPSRL"
1313..1325
BASE COUNT 314 a 333 c 407 g 271 t
ORIGIN
alignment_scores:
Quality: 1399.00 Length: 278
Ratio: 5.032 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-866-034-2 x AB045131
Align seq 1/1 to: AB045131 from: 1 to: 1325
1 MethIlsysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
80 ATGCACAGGCGGGGCTGTGAGGCTCTGTGCGCGGGCTTGGAATTCGGT 129
17 lArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
130 GCGGATGGCCAGCTCCGGGATGACCCGCGGACCGGCTGCAAAATAGG 179
34 alAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
180 TGGCCCTGTGTAACGCGCTCCACCGAGCGGATCGCTTCGCATCGCCGG 229
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnG 67
230 CGTTTGGCCGAGGACGGGGCCCATGTGTGTCACAGCGGACGCAACGCA 279
67 nasnValaAspGlnAlaValAlaThrLeuGlnGlnGlyGluLeuSerVal 84
280 GAATGTGACACAGCGCGTGCACCGTGCAGGGGGAGGGCTGAGGCTGA 329
84 hGlyThrValCysHisValGlyLysAlaGlnAspArgLysArgLeuVal 100
330 CGGCGACCGTGTGCCATGTGGGGAAGCGGAGACCGGCGGCTGTGG 379
101 AlathAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
380 GCCACGCGTGTGAAGCTCATGTGATGATCTCTGATCTCCAAATGC 429
117 alAlaValaAsnProPhePheGlySerIleMetAspValThrGluGluVal 134
430 TGCTGTCAACCCCTTCTTGGAAAGCATTAATGATGATGATGAGAGCTGT 479
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
480 GGGACAGAGCTGTGACATTAAATGAGAGCCGCCAGCTGATGACAAAG 529
151 AlalaValaProglumetglulysarglyglySerValaIleVa 167
530 GCAGTGTGTCAGAAATGAGAAACAGAGAGCGGCTCAGTGTGATCGT 579
167 lSerSerIleAlaIleAlaPheSerProSerProgluPheSerProtyrAsn 184
580 GTCTTCATAGCAGCCTTCATGCTCATCTCTGCTGCTCAGTCTTACAAG 629
184 alSerlysthAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
630 TCAGTAAACAGCCTTGTGGGCTGACCAAGACCTGCGCATAGAGCTG 679
201 AlaproArgAsnIleArgValaAsnCysLeuAlaProgluLeuIleLysTh 217

```



```

251 11eValSerPheLeuCySerGluAspAlaSerTyrIleThrGlyLue 267
|||||
771 ATCGTGTCTTCTCTGCTCTGAAGATGCCAGTACATCATCGGGGAAC 820
267 rValValValGlyGlyThrProSerArgLeu 278
|||||
821 AGTGGTGGTGGTGGAGAACCCCGTCCGCCCTC 854

seq_name: gb_AF044127
seq_documentation_block:
LOCUS AF044127 1281 bp mRNA linear PRI 27-MAY-1999
DEFINITION Homo sapiens peroxisomal short-chain alcohol dehydrogenase
(SCAD-SRL) mRNA, complete cds.
ACCESSION AF044127
VERSION AF044127.1 GI:4105189
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1281)
AUTHORS Franssen, M., Van Veldhoven, P. P. and Subramani, S.
TITLE Identification of peroxisomal proteins by using M13 phage protein
VI phage display: molecular evidence that mammalian peroxisomes
contain a 2,4-dienoyl-CoA reductase
Biochem. J. 340 (Pt 2), 561-568 (1999)
JOURNAL 9926733
MEDLINE 2 (bases 1 to 1281)
REFERENCE Franssen, M. and Subramani, S.
AUTHORS Direct Submission
TITLE Submitted (22-JAN-1998) Biology, University of California, San
JOURNAL Diego, Gilman Drive, La Jolla, CA 92093-0322, USA
FEATURES
Source
1..1281
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1281
/gene="SCAD-SRL"
misc_feature
1..1281
/gene="SCAD-SRL"
/note="Similar to ESTs with GenBank Accession Numbers
R25850, AA293652 and H63672"
80..862
/gene="SCAD-SRL"
/note="SCAD-SRL; contains a C-terminal peroxisomal
targeting signal (PTS1)"
/codon_start=1
/product="peroxisomal short-chain alcohol dehydrogenase"
/db_xref="GI:4105190"
/translaton="MASSGMRTRDPLANKVALVYASTDGIQFAIARLADGAHVVS
SRQQNVDAVATLOEGLSVTGVCHVGAEDREERLVAAVKLHGIDILVNAAVN
PFGSLMDVTEYWDKTDINVKAPALMTKAVVPEMKRGGSVIVISSTIAFSPSPG
FSFYNSKTKALIGTKTLATELAPRINRVNCLAPGLIKTSPESRLMMDKEESMKET
LRIRLRGEPPDCAGIYSFLCSEASITTTGFTVVVGGTSPSRL"
polyA_signal
1235..1240
/gene="SCAD-SRL"
BASE COUNT 300 a 323 c 395 g 263 t
ORIGIN
alignment_scores:
Quality: 1391.00 Length: 278
Ratio: 5.022 Gaps: 0
Percent Similarity: 99.640 Percent Identity: 99.281
alignment_block:
US-09-866-034-2 x AF044127 ..
Align seg 1/1 to: AF044127 from: 1 to: 1281
1 MethiSLySAlaGlyLeuLeuGlyLeuCySAlaArGaIaTPrAsnSerVa 17

```

```

26 ATGCACAGGCGGGGCTGTAGGCTCTTGCCCGGGCTTGGAATTCGGT 75
17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
76 GCGGATGGCAGCTCCGGATGACCCCGCGGACCCGCTGCAAAATTAAG 125
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
126 TGGCCCTGGTAAACGGCTCCACCGACGGATCGGATTCGCCATCGCCG 175
51 ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgGlnGln 67
176 CGTTGGCCAGAGCGGGCCCATGTGTGTGACAGCCGAGGAGACGCA 225
67 naSValAspGlnAlaValAlaThrLeuGlnGlnGlyGlyLeuSerValT 84
226 GATGTGGACACAGCGCGTGGCCACGCTGCAGGGGAGGGGCTAGCGCTGA 275
84 hGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
276 CCGGACCCGTGTGCCATGTGGAGAGCGGAGACCGGAGCGGTGGTG 325
101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
326 GCATGGCTGTGAAGCTTCATGGAGGTATCGATATCTCATGCAATGC 375
117 aAlaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
376 TGCTGTCAACCTTCTTCTTGGAAGCCTAATGATGATCATGACGAGGTGT 425
134 rPaSpLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
426 GGGACAAGACTGTGACATTAATGTGAAGGCCACCCGATGACACAAG 475
151 AlaValValProGluMetGluLysArgGlyGlyGlySerValValIleVa 167
476 GCGGTGTGCCAGAAATGGAGAACAGAGGCGGCTCATGTGTCATCGT 525
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
526 GTCTTCATAGCAGCCTTCATGTCATCTCTGCTTCACTTACAAAG 575
184 aISerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
576 TCAGTAAACAGCCTTGCTGGGCTGACCAAGACCTTGCCATAGAGCTG 625
201 AlaProArgAsnIleArgValaCysLeuAlaProGlyLeuIleLysTh 217
626 GCGCCAGGAACATTAGGGTGAACCTCCATGACCTGGAATTATCAAGAC 675
217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysMetL 234
676 TACCTTCACAGATCTCTGATGACAAAGAAAGAGGAAAGATGATGA 725
234 ySGLuThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGly 250
726 AAGAAACCTCGCGATAGAGAGGTAGGGGAGGACAGAGATTGTGTGGC 775
251 11eValSerPheLeuCySerGluAspAlaSerTyrIleThrGlyLue 267
|||||
776 ATCGTGTCTTCTCTGCTCTGAAGATGCCAGTACATCATCGGGGAAC 825
267 rValValValGlyGlyThrProSerArgLeu 278
|||||
826 AGTGGTGGTGGTGGAGAACCCCGTCCGCCCTC 859

seq_name: gb_cm:AB062757
seq_documentation_block:
LOCUS AB062757 880 bp mRNA linear MAM 04-DEC-2001
DEFINITION Sus scrofa mRNA for carbonyl reductase/NADP-retinol
dehydrogenase,
complete cds.

```

ACCESSION AB062757
 VERSION AB062757.1 GI:17298118
 KEYWORDS
 SOURCE Sus scrofa male heart cDNA to mRNA.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 REFERENCE
 AUTHORS Hara, A., Imamura, Y., Abe, H., Inoue, S. and Ishikura, S.
 TITLE Molecular characterization of a carbonyl/retinal reductase of pig tissues
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 880)
 AUTHORS Hara, A.
 JOURNAL Direct Submission
 Submitted (06-JUN-2001) Akira Hara, Gifu Pharmaceutical University, Laboratory of Biochemistry, Mitahara-higashi 5-6-1, Gifu, Gifu 502-8585, Japan (E-mail: hara@ifu-pu.ac.jp, Tel: 81-58-237-8586, Fax: 81-58-237-8586).
 FEATURES
 source location/Qualifiers
 1..880
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /sex="male"
 /tissue_type="heart"
 28..810
 /EC_number="1.1.1.184"
 /codon_start=-1
 /product="carbonyl reductase/NADP-retinol dehydrogenase"
 /protein_id="BAB78528.1"
 /db_xref="GI:17298119"
 /translation="MASTGVERRRPLENKVALVTASTDGILAIARLADGAAHYVS SRKQNDVTATLGEGLSYCTGVCHGKAEKREPLVAAVNLHGGVDILVSNALVN PFGNTIIDTEYWKDILVNNKATVLMTRAVVPEKKRGSGVLLVSVGAHPFN LQRYNSKTALGLTKNLVAELRVNCLAPGKIKTFNSQVLMMDKARKREYRES LRRLRGNEPDAGIVSFLCSEDAVSITGETVYVGGGTASRL"
 BASE COUNT 205 a 224 c 288 g 163 t
 ORIGIN
 alignment_scores:
 Quality: 1154.00 Length: 269
 Ratio: 4.490 Gaps: 0
 Percent Similarity: 95.539 Percent Identity: 83.271
 alignment_block:
 US-09-866-034-2 x AB062757 ..
 Align seg 1/1 to: AB062757 from: 1 to: 880
 10 CysAlaIArgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrAr 26
 1 TGTTCACAGACGTGTAGTGGTGGGACGACCGGGGGGAGCG 50
 26 GATGASPProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspG 43
 51 CCGAAACCGCTCGAGATAGTGCGACTGTAACAGCCCTCCACTGACG 100
 43 LylleGlyPheAlaIleAlaIArgArgLeuAlaGlnAspGlyAlaHisVal 59
 101 GGATCGGCTCTCCATCGCCCGCTGCGCCAGATGGGGCCACAGTG 150
 60 ValValSerSerArgLysGlnGlnAsnValAspLlnAlaValAlaThrle 76
 151 GTGCTCAGCACCGCGAAGCAGAGATGTGACCGACAGTGGCAACACT 200
 76 uGlnGlyLugLylLeuSerValThrGlyThrValCysHisValGlyLysA 93
 201 GCAAGGGGAGGGCGCTGAGCGGAGCGGACCGCTGCCACGCTGGGGAAG 250
 93 laGluAspArgGluArgLeuValAlaThrAlaValLysLeuHisGlyCly 109
 251 CCGAGGACCGGGAGCGACTGCTGCGCTGTGAACCTTCACGGGGGC 300

110 IleAspIleLeuValSerAsnAlaAlaValAsnProPheGlySerI 126
 301 GTGGACATCTGCTGCTCCATGACAGCCGTCACCATCTTGTGAAACAT 350
 126 eMetAspValThrGluGluValTrpAspLysThrLeuAspIleAsnVal 143
 351 CATGATGCCACTGAGGAGGTGGGACAAAGTTCCTGCACTGTAACGCA 400
 143 ySalAlaProAlaLeuMetThrLysAlaValAlaProGluMetGluLysArg 159
 401 AGCCACAGCTCTCTGATGACAAAGCGAGTGTCCCGAGATGGAGAACCA 450
 160 GlyGlyGlySerValIleValSerSerIleAlaAlaPheSerProse 176
 451 GGAGGCGGCTCAGTGTGATCTATCCCTCCGTTGGAGCCTACCAACCAAT 500
 176 rProGlyPheSerProTyrAsnValSerLysThrAlaLeuLeuGlyLeu 193
 501 TCCCATCTGGGCTTCAATGTTAGTAAACACCTTGTGCTGGGCTCA 550
 193 hTlyThrLeuAlaIleGluLeuAlaProArgAsnIleArgValAsnGly 209
 551 CCANAGACCTGCGCCCTAGAGCTGCGCCCAAGAACATTAGGTGAACCTGC 600
 210 LeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeuTrpMetAs 226
 601 CTGGGCGCTGGACATCATCAGACTTAACCTGAGGAGGTGTGTGATGGA 650
 226 pLysGluLysGluGluSerMetLysGluThrLeuArgIleArgGluLeu 243
 651 TAAGGCAAGAAAGAGTACATGAAGAAAGCTGCAATCAAGAAAGTTAG 700
 243 lYgluProGluAspCysAlaGlyIleValSerPheLeuGlySerGluAsp 259
 701 GCAATCCAGAGAGACTGCGGCGCATCGTCTTCTTCTTCTTCTTCTTAAAT 750
 260 AlaSerTyrIleThrGlyGluThrValValGlyGlyGlyThrProse 276
 751 GCCAGCTTACATCAGTGGGAGACAGTGTAGTGGTGGAGGAGCGCATC 800
 276 rArgLeu 278
 801 CCGCCTC 807
 seq_name: gb_ro:AB045132
 seq_documentation_block:
 LOCUS AB045132 1312 bp mRNA linear ROD 05-DEC-2000
 DEFINITION Mus musculus mounRDR mRNA for NADPH-dependent retinol dehydrogenase/reductase, complete cds.
 ACCESSION AB045132
 VERSION AB045132.1 GI:11559413
 KEYWORDS
 SOURCE
 ORGANISM Mus musculus (strain:C57 black) liver cDNA to mRNA.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Furukawa, A., Ohnishi, T., Huang, D., Araki, N. and Ichikawa, Y.
 TITLE CDNA cloning and characterization of peroxisomal short-chain dehydrogenase / reductase that reduce all-trans retinal to retinol
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1312)
 AUTHORS Furukawa, A., Huang, D., Ohnishi, T. and Ichikawa, Y.
 JOURNAL Direct Submission
 Submitted (23-JUN-2000) Aizo Furukawa, Kagawa Medical University, Faculty of Medicine, Department of Biochemistry, Miki-cho, Ikenobe 1750-1, Kita-gun, Kagawa 761-0793, Japan
 (E-mail: aifurukawa@kms.ac.jp; Tel: 81-87-891-2104, Fax: 81-87-891-2105)
 FEATURES
 source location/Qualifiers
 1..1312
 /organism="Mus musculus"

```

/strain="C57 black"
/db_xref="taxon:10090"
/tissue_type="Liver"
115..897
/gene="mounRDR"
115..897
CDS
/gene="mounRDR"
/codon_start=1
/product="NADPH-dependent retinol dehydrogenase/reductase"
/db_xref="GI:11559416"
/translation="MASSGCTRRNPISNKVALYATASDGIQFALARRLEAGAHVYVS
SRKOQNDRAVATITOGESLVTGIVCHGKRADEKLTITLKRHOGLDILVNAVN
PFGNLMQVTEVWDKVLSTINVTATAMIKAVPEMRGGGSVIVGSAVGTFRPS
LGPVNSKTLVLGTRNFAELPKNIRVNCIALPKIKTRFSVLMEEKAREDEIKKA
MOIRLRKPEDCAGIVSFLCSEDASYINGETVVVGGGPPSRL"
POLYA_site
1303..1312
BASE COUNT 328 a 305 c 385 g 294 t
ORIGIN

alignment_scores:
Quality: 1130.50 Length: 279
Ratio: 4.331 Gaps: 1
Percent Similarity: 93.548 Percent Identity: 80.287

alignment_block:
US-09-866-034-2 x AB045132

Align seg 1/1 to: AB045132 from: 1 to: 1312

1 MethiSLysAlaGly...LeuLeuGlyLeuCySAlaArgAlaTrpAsnSe 16
|||||
58 ATGCAGAAAGCGGACGGCTGTAGTGGTGTGACTCAACGCGTGCATGTC 107
16 rValArgMetAlaSerSerglyMethrArgArgAspProLeuAlaAsnL 33
|||||
108 GGTGGAGATGGCCAGTTCGGGTGACTGCTGCAACCTCTCTCCAAATTA 157
33 yValAlaLeuValThrAlaSerThrAspGlyLleGlyPheAlaIleAla 49
|||||
158 AGGTGGCCCTGTGCACAGCTCCACCGAGGATGGCTTGCCTGCATCGCC 207
50 ArgArgLeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysL 66
|||||
208 CGGCGCGCTGGTGAAGATGGGGCCACGTCGTCTCAGACCGCAACA 257
66 nGlnAsnValAspGlnAlaValAlaThrLeuGlnGlyLysLysLeuSerV 83
|||||
258 GCAGATGTGCACCTGCGACAGTGCACACTACAGGAGAGGCGCTGAGTC 307
83 aThrGlyThrValCysHisValGlyLysAlaGlnAspArgGluArgLeu 99
|||||
308 TCACGTGGCATCTGTGCCACGTGGGGAAGGAGAGATGAGAAAAAGCTG 357
100 ValAlaThrAlaValLysLeuHisGlyLysLysLysLysLysLysLys 116
|||||
358 ATACACACCGCTGTGACAGCTCACCGAGGATGATATCTGTCTCCAA 407
116 nAlaAlaValAsnProPhePheGlySerLleMetAspValThrGluLys 133
|||||
408 TGCTCTGTCAACCTTCTTGGAAATCTAATGATGTGCACAGAGAGAG 457
133 aLTrpAspLysThrLeuAspLysValAlaProAlaLeuMetThr 149
|||||
458 TGTGGACAAAGGTTTGAACATTATGTGACAGCTACACCCATGATGATT 507
150 LysAlaValAlaProGluMetGluLysArgGlyLysLysValValAla 166
|||||
508 AAGCGGTGTGTCCAGATGTGAAAAAGCAGAGGCGGCTCAGTGTGAT 557
166 eValSerSerIleAlaIlePheSerProSerProGlyPheSerProTyr 183
|||||

```

```

558 TGTGGCTTCTGTAGCAGGCTTCACGTGCTCCCTCTGCGGCTCTTACA 607
183 snValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 199
|||||
608 ATCTTGACCAACACCTTTGCTGTCTTACTAGAACCTTTGACGGGAG 657
200 LeuAlaProArgAsnIleArgValAsnCySLeuAlaProGlyLeuLys 216
|||||
658 TTGGCCCGGAGAACATTCAGAGTACGCTGTGACACCTGACTCATTCAA 707
216 sThrSerPheSerArgMetLeuTrpMetAspLysGluLysGluSerM 233
|||||
708 GACTCATTCACGACATCTCTTGTGGAGAGGAAGACAGAGACTTCA 757
233 eLysGluThrLeuArgGlyLeuArgArgLeuGlyLysProGluAspCySAla 249
|||||
758 TAAAGAGCGCATGCAAAATCAGAGGCTAGGACGACGAGATTTGTGCT 807
250 GlyIleValSerPheLeuCySArgLysAspAlaSerTyrIleThrGlyL 266
|||||
808 GCGATGTTTCCCTTCTTATGCTCTGAAAGCCGACGTTACATCATGAGA 857
266 uThrValValAlaGlyGlyThrProSerArgLeu 278
|||||
858 GACCGTAGTAGTGGGGGAGAACCCCTTCTCGCCTC 894

seq_name: gb_com:AB045133

seq_documentation_block:
LOCUS AB045133 1212 bp mRNA linear MAM 05-DEC-2000
DEFINITION Oryctolagus cuniculus rabnRDR mRNA for NADPH-dependent retinol
dehydrogenase/reductase, complete cds.
ACCESSION AB045133
VERSION AB045133.1 GI:11559415
KEYWORDS
SOURCE Oryctolagus cuniculus liver cDNA to mRNA.
ORGANISM Oryctolagus cuniculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE
1 (sites)
Furukawa,A., Ohnishi,T., Huang,D., Araki,N. and Ichikawa,Y.
cDNA cloning and characterization of peroxisomal short-chain
dehydrogenase / reductase that reduce all-trans retinal to retinol
unpublished
2 (bases 1 to 1212)
Furukawa,A., Huang,D., Ohnishi,T. and Ichikawa,Y.
Direct Submission
Submitted (23-JUN-2000) Atzo Furukawa, Kagawa Medical University,
Faculty of Medicine, Department of Biochemistry; Miki-cho, Ikenobe
1750-1, Kita-gun, Kagawa 761-0793, Japan
(E-mail:afurukawa@kms.ac.jp, Tel:81-87-891-2104, Fax:81-87-891-2105)
FEATURES
source
1..1212
/organism="Oryctolagus cuniculus"
/db_xref="taxon:9986"
/tissue_type="Liver"
1..783
/gene="rabinRDR"
1..783
CDS
1..783
/gene="rabinRDR"
/codon_start=1
/product="NADPH-dependent retinol dehydrogenase/reductase"
/protein_id="BAB1877.1"
/db_xref="GI:11559416"
/translation="MASSGVTRRDPLANKVAIVTASDGIQFALARRLAODGAAHVYS
SRKOQNDRAVATITOGESLVTGIVCHGKRADEKLTITLKRHOGLDILVNAVN
PFGNLMQVTEVWDKVLSTINVTATAMIKAVPEMRGGGSVIVGSAVGTFRPS
LGPVNSKTLVLGTRNFAELPKNIRVNCIALPKIKTRFSVLMEEKAREDEIKKA
MOIRLRKPEDCAGIVSFLCSEDASYINGETVVVGGGPPSRL"
POLYA_site
1203..1212
BASE COUNT 268 a 336 c 375 g 233 t
ORIGIN

```

alignment_scores:
 Quality: 1123.00 Length: 260
 Ratio: 4.492 Gaps: 0
 Percent Similarity: 96.154 Percent Identity: 85.385

alignment_block:
 US-09-866-034-2 x AB045133

Align seg 1/1 to: AB045133 from: 1 to: 1212

```

19 MetLaserSerGlyMetThrArgArgAspProLeuAlaAsnLysValAl
1 ATGCCAGACTCTGGGTGATACCCCGGAGACCCGCTCGGCAFAAGGTGCG
35 aleuValThrLaserThrAspGlyIleGlyPheAlaIleAlaArgAyl
1 :|||||
51 CATGTGATGCGCTCCACCGAGGATGCTGGCCATCCCGCGGCTC
52 euAlaGlnAspGlyAlaHisValValValSerSerArgLysGlnGlnAsn
101 TGGCCAGAGATGGGGCCACCTGCTCATACAGCCGAGACAGAGAC
69 ValAspGlnAlaValAlaThrLeuGlnGlyGlnGlyLeuSerValThrG
151 GTGGACCGCGCGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
85 YThrValCysHisValGlyLysAlaGlnAspArgGlnArgLeuValAla
201 CACGGTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
102 hValAlaValLysLeuHisGlyGlyLysAlaGlnAspArgGlnArgLeu
251 CGGCGCTGAGACCTCCACGAGGATGATATCTGCTGCTGCTGCTGCTGCTG
119 ValAsnProPhePhePheGlySerIleMetAspValThrGlnGlnVal
301 GTCAACCCCTCTTTGGAAAGCTATGATGATGATGATGATGATGATGATG
135 pLysThrLeuAspLysAlaValLysAlaProAlaLeuMetThrLysAla
351 CAGATCTTGACATTAACTGAAAGCCATGCTGCTGCTGCTGCTGCTGCTG
152 alValProGlnMetGlnLysArgGlyLysValSerValIleValSer
401 TGGTGCCAGAGATGAGAAACGAGAGCGGCTCGGTGCTGCTGCTGCTG
169 SerIleAlaAlaPheSerProSerProGlyPheSerProGlyAsnVal
451 TCCATAGCAGCCCTTCAATCATTGAGTGGCTGGGCTGCTGCTGCTGCTG
185 rLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlnLeuAla
501 TAAACAGCCCTTGGTGGCTTACCAAGAACCTGCTGCTGCTGCTGCTG
202 roArgAsnIleArgValAsnLysAlaValProGlyLeuIleLysThrSer
551 CCAGAGACATCCGGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
219 PheSerArgMetLeuThrMetAspLysGlnLysGlnLysSerMetLys
601 TTCAGTAGGCGTGTGGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAG
235 uThrLeuArgIleArgArgLeuGlyGlnProGlnAspCysAlaGlyIle
651 AAGGTGAGATAGAGAGAGGTAGCAAAACAGAGAGATGCTGCGGCAATG
252 alSerPheLeuCysSerGlnAspAlaSerGlyThrArgLysGlnThrVal
701 TGTCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
269 ValValGlyLysGlyThrProSerArgLeu 278
|||||

```

751 GGTGTGGCTGGAGAGACCACTCCCGCTC 780

seq_name: gb.ro:BC003484

seq_documentation_block:
 LOCUS BC003484 1234 bp mRNA linear ROD 12-JUL-2001
 DEFINITION Mus musculus, similar to peroxisomal short-chain alcohol
 dehydrogenase, clone MGC:6560 IMAGE:2811468, mRNA, complete cds.
 ACCESSION BC003484
 VERSION BC003484.1 GI:13097509
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1234)
 TITLE Strausberg, R.
 JOURNAL Direct Submission
 Submitted (20-FEB-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalobos@bcm.tmc.edu
 Villalobos, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 5 Row: n Column: 8.
 Location/Qualifiers
 1. 1234
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:6560 IMAGE:2811468"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
 old, gross tissue."
 /clone_lib="NCI CGAP_Mam5"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 38..820
 /codon_start=1
 /product="similar to peroxisomal short-chain alcohol
 dehydrogenase"
 /protein_id="AAH03484.1"
 /db_xref="GI:13097510"
 /translation="MASSGLTFRNP LSKNVALVTA STDGI GAIRARLAEAGAHVYS
 SRKQONVRAVAATLOGEGLSVTGIVCHVGRKEDREKLITLTKLRHGGIDIIYSNAAYV
 PFGNLMVTEVEMDKVLSINVTATRAMIKAVPEMEKRGSGSVYIVGSAVGFTRFPS
 LCPYVSTALILGLTKNRAELAPKAVINVCAPLITRFSYVIMEKRAEDFLKEA
 MOIRLGRPEDDAGIVSFLCSDDAIVNGETVYVVGGRPSRL"

BASE COUNT 313 a 273 c 367 g 281 t

ORIGIN

alignment_scores:
 Quality: 1117.00 Length: 272
 Ratio: 4.380 Gaps: 0
 Percent Similarity: 93.750 Percent Identity: 80.515

alignment_block:
 US-09-866-034-2 x BC003484

Align seg 1/1 to: BC003484 from: 1 to: 1234

```

7 LeuGlyLeuCysAlaArgAlaTrpAsnSerValArgMetAlaSerSerg 23
  ||||| :||| ||||| ||||| ||||| |||||
2 TTAGGTGGCTGGACGACGATGATGCTGGTGGATGGCCGCTCCG 51
  ||||| ||||| ||||| ||||| ||||| |||||
23 yMetThrArgArgAspProLeuAlaAsnValAlaLeuValThrAla 40
  ||||| ||||| ||||| ||||| ||||| |||||
52 GTTGACTCTCCGAACCTCTCTCGAATAAGTGGCTGGTCACAGCCT 101
  ||||| ||||| ||||| ||||| ||||| |||||
40 exThraspGlyIleGlyPheAlaIleAlaArgArgLeuAlaGlnaspGly 56
  ||||| ||||| ||||| ||||| ||||| |||||
102 CCACGACGAGGATGGCTTGGCATGCCCCGCTGCTGGCTGAAGATGG 151
  ||||| ||||| ||||| ||||| ||||| |||||
57 AlaHisValValSerSerArgGlyGlnAsnValAlaSpGlnAlaVal 73
  ||||| ||||| ||||| ||||| ||||| |||||
152 GCCACAGCTGGTGTGTCAGCAGCGCAACAGCAAGATGTGACCGTCA 201
  ||||| ||||| ||||| ||||| ||||| |||||
73 AlaThrLeuGlnGlyGlyLeuSerValThrGlyThrValCysHisVal 90
  ||||| ||||| ||||| ||||| ||||| |||||
202 GCCACACTACAGAGGAGGAGGCTGAGTGTGATGTCATGCTGGTGCAC 251
  ||||| ||||| ||||| ||||| ||||| |||||
90 aGlyValAlaGluAspArgGluArgLeuValAlaThrAlaValLysLeu 106
  ||||| ||||| ||||| ||||| ||||| |||||
252 TGGGGAAGGACAGAGGATCGAGAAAGCTGATACACAGGCTGTGAAG 301
  ||||| ||||| ||||| ||||| ||||| |||||
107 HisGlyGlyIleAspIleLeuValSerAsnAlaAlaValAspProPhe 123
  ||||| ||||| ||||| ||||| ||||| |||||
302 CACCGGGGATGATGATATCTGCTCCATGCTGCTGATACCCCTTTCT 351
  ||||| ||||| ||||| ||||| ||||| |||||
123 eGlySerIleMetAspValThrGluGluValTrpAspLysThrLeuAsp 140
  ||||| ||||| ||||| ||||| ||||| |||||
352 TGGAAATCTAAATGATGTCTCACAGAGAGGTGTGGGACAAAGTTTGA 401
  ||||| ||||| ||||| ||||| ||||| |||||
140 LeasValAlaValProAlaLeuMetThrLysAlaValAlaProGluMet 156
  ||||| ||||| ||||| ||||| ||||| |||||
402 TTAATGTGACACTACAGCATGATGATTAAGCGGTGTGTCAGAGATG 451
  ||||| ||||| ||||| ||||| ||||| |||||
157 GluLysArgGlyGlyLysSerValValIleValSerSerIleAlaAla 173
  ||||| ||||| ||||| ||||| ||||| |||||
452 GAAAGAGCGAGGAGCGGCTCAGTGTGATGTGGGTCTGTGACAGGCT 501
  ||||| ||||| ||||| ||||| ||||| |||||
173 eSerProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeu 190
  ||||| ||||| ||||| ||||| ||||| |||||
502 CACTGGTTCCTCTCTGGGTCTTACAAATGTAGCAAAACAGCTTTC 551
  ||||| ||||| ||||| ||||| ||||| |||||
190 euGlyLeuThrLysThrLeuAlaIleGluLeuAlaProArgAsnIleArg 206
  ||||| ||||| ||||| ||||| ||||| |||||
552 TGGGCTTACTAAGAACTTTGACAGCGGAGTTGGCCCGAAGAACATTC 601
  ||||| ||||| ||||| ||||| ||||| |||||
207 ValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLe 223
  ||||| ||||| ||||| ||||| ||||| |||||
602 GTGAAGCTGTAGCACCTGAGCTCATCAAGACTCGATTCAGCAGTGTCT 651
  ||||| ||||| ||||| ||||| ||||| |||||
223 uTrpMetAspLysGlyLysGluLysSerMetLysGluThrLeuArgIleAla 240
  ||||| ||||| ||||| ||||| ||||| |||||
652 GTGGAGAGGAAAGCAAGAGAGACTTCAATAAAGAGCATGCAAAATCA 701
  ||||| ||||| ||||| ||||| ||||| |||||
240 rGArgLeuGlyLysProGluAspCysAlaGlyIleValSerPheLeuGly 256
  ||||| ||||| ||||| ||||| ||||| |||||
702 GAAGGCTAGGCAAGCAGAGATTTGTGCTGGCATAGTTCTCTTATTC 751
  ||||| ||||| ||||| ||||| ||||| |||||
752 TCTGAAGAGCCAGTTACATCAATGAGAGACCGTAGTAGTGGGGAGAG 801
  ||||| ||||| ||||| ||||| ||||| |||||
273 yThrProSerArgLeu 278
  ||||| ||||| ||||| ||||| ||||| |||||
802 AACCCCTTCTGCGCTC 817
  ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_ro:AB062758
seq_documentation_block:

```

```

LOCUS      AB062758                806 bp    mRNA    linear    ROD 04-DEC-2001
DEFINITION Rattus norvegicus mRNA for carbonyl reductase/NADP-retinol
            dehydrogenase, complete cds.
ACCESSION  AB062758
VERSION    AB062758.1 GI:17298120
KEYWORDS   Rattus norvegicus (strain:Wistar) male heart cDNA to mRNA.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (sites)
AUTHORS   Hara,A., Imanura,Y., Abe,H., Inoue,S. and Ishikura,S.
TITLE     Molecular characterization of mammalian carbonyl/retnal reductases
JOURNAL   Unpublished
AUTHORS   2 (bases 1 to 806)
Hara,A.
FEATURES   Direct Submission
            Submitted (06-JUN-2001) Akira Hara, Gifu Pharmaceutical University,
            Laboratory of Biochemistry, Mitahora-higashi 5-6-1, Gifu, Gifu
            502-8585, Japan (E-mail:hara@ifu-pu.ac.jp, Tel:81-58-237-8586,
            Fax:81-58-237-8586)
            Location/Qualifiers
                source          1..806
                                /organism="Rattus norvegicus"
                                /strain="Wistar"
                                /db_xref="taxon:10116"
                                /sex="male"
                                /tissue_type="heart"
                                /EC_number="1.1.1.184"
                                /codon_start=1
                                /product="carbonyl reductase/NADP-retinol dehydrogenase"
                                /protein_id="BAB78529.1"
                                /db_xref="GI:17298121"
                                /translation="MASSGLTRQNPANLKVATSTDGIGLAIRRLAEDGAHVVIS
                                SRKOQNDRAVALTQEGELSVTVGVCHVKADEDEKLIYNNALKIHOGIDILVSAAVN
                                PFGNLMQDVEEVNNKVLSTINVTASAMIKAVVPAEMKRGGSVTVISVAVPLFPS
                                LGPNVNSTALIGITKNPAELAKNIRVNCRLAAGLIKTRHSSVLMKEAREMKEKET
                                MOIRRLKPEPCVGSIVSFLESDSYINGETVVGGGTPSHL"
                CDS              202 a 187 c 240 g 177 t
                BASE COUNT      202 a 187 c 240 g 177 t
                ORIGIN
alignment_scores:
            Quality: 1073.00      Length: 260
            Ratio: 4.398          Gaps: 0
            Percent Similarity: 93.846      Percent Identity: 81.154
alignment_block:
US-09-866-034-2 x AB062758
Align seg 1/1 to: AB062758 from: 1 to: 806
19 MetAlaSerSergIyMetThrArgArgAspProLeuAlaAsnValAla 35
  ||||| ||||| ||||| ||||| ||||| |||||
1 ATGGCGACGTTCGGGGCTGACTGCGCCAGAACCCACTGCCAATTAAGTGGC 50
  ||||| ||||| ||||| ||||| ||||| |||||
35 AleuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArgL 52
  ||||| ||||| ||||| ||||| ||||| |||||
51 CCTGGTCACAGCCTCCACGAGGAGTTGGCTTGGCATGCGCCGCGCGTC 100
  ||||| ||||| ||||| ||||| ||||| |||||
52 euAlaGlnaspGlyAlaHisValValSerSerArgLysGlnAsn 68
  ||||| ||||| ||||| ||||| ||||| |||||
101 TGGCTGAGGAGCGGGCCGCTGCTATCAGCAGCGCAAGAGCAAGT 150
  ||||| ||||| ||||| ||||| ||||| |||||
69 ValAspGlnAlaValAlaThrLeuGlnGlyGlyLysSerValThrGly 85
  ||||| ||||| ||||| ||||| ||||| |||||
151 GTGGACCGTGGCAGTGGCCACACTCCAGGAGAGGAGCTGATGTTACTGG 200
  ||||| ||||| ||||| ||||| ||||| |||||
85 yThrValCysHisValGlyLysAlaGluAspArgGluArgLeuValAla 102
  ||||| ||||| ||||| ||||| ||||| |||||
201 CGTGTGTGTCATGTGGGGAAGCAGAGAGATGAGAAAGCTGTAAACA 250
  ||||| ||||| ||||| ||||| ||||| |||||

```

```

102 hralaValLysLeuHISglyIleAspIleLeuValSerAsnAla 118
119 ValAspProPhePheGlySerIleMetAspValThrGluIuValTyrAs 135
301 GTCACACCTTTCTTTGGAAATCTAATGACGCTCAGACAGAGAGGTGGAA 350
135 pLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAla 152
351 CAAGGTTTGAGCATTAATGTGACAGCTTCAGCAATGATTAAGCGG 400
152 aLValProGluMetGluLysArgGlyGlySerValIleValSer 168
401 TGGTGGCAGCATGAAACAGAGAGGCGGCTCGGTGATGTGAGT 450
169 SerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnValSe 185
451 TCCGTAGCAGGCTTCCTTATTCCTCTGCGCCCTTACAAATGTAG 500
185 rLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAla 202
501 CAACACAGCTTCTGCTGCTCAGTAAAGACTTGCAGAGTGGCC 550
202 rArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSer 218
551 CAAGAACATTCGAGTGAAGCTGCTGCGACCTGACTCATCAGACTCAC 600
219 PheSerArgMetLeuThrMetAspLysGlyLysGluGluSerMetLysG 235
601 TTCAGCAGTGTGTGTGGAAGCAAGACAGAAATGATTAAGA 650
235 uThrLeuArgIleArgArgLeuGlyLysGluProGluAspCysAlaGlyIle 252
651 AACCATGCAATTTAGAGCTTAGCAGCAGAGGATGTGTGCTGATAG 700
252 aLSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrVal 268
701 TTTCCTTCTTGTGCTCTGAAGATGCCAGTTACATCAATGCGAGACAGT 750
269 ValValGlyGlyGlyThrProSerArgLeu 278
751 GTGCTTGGGGGAGAACCCCTTCTCCCTC 780

seq_name: gb_pr:AY071856

seq_documentation_block:
LOCUS AY071856 1003 bp mRNA linear PRI 20-JAN-2002
DEFINITION Homo sapiens NADP-dependent retinol dehydrogenase short isoform
ACCESSION AY071856
VERSION AY071856.1 GI:18252126
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 1003)
AUTHORS Du,J., Liu,G.F., Wang,B., Sun,T., Chang,Y.M., Vennuste,G., Zhu,L.,
Chen,K.M., Wang,G.L., Xu,X.L. and Huang,D.Y.
Cloning and characterization of short NADPH-dependent retinol
dehydrogenase/reductase that reduces all-trans retinal to retinol
Unpublished
2 (bases 1 to 1003)
JOURNAL Du,J., Liu,G.F., Wang,B., Sun,T., Chang,Y.M., Vennuste,G., Zhu,L.,
REFERENCE Chen,K.M., Wang,G.L., Xu,X.L. and Huang,D.Y.
AUTHORS Direct Submission
TITLE Submitted (26-DEC-2001) Cell Biology, China Medical University,
JOURNAL No.92 Beier Road,Heping District, Shenyang, Liaoning 110001, China
FEATURES
SOURCE 1..1003
/organism="Homo sapiens"

```

```

CDS
<
/db_xref="taxon:9606"
6..584
/codon_start=1
/product="NADP-dependent retinol dehydrogenase short
isoform"
/protein_id="AAL61824.1"
/db_xref="GI:18252127"
/translation="MHKAGLGLCARANNSSGMTTRDPLANKVALVYASTDGI
GFAIRLAQDGAHVVSRRQONDOAVATLQCGLSVTGVCVKAEDRERLVAT
AVKHGIDILVSNAAVNPFEGLINDVIEVWDKLMDEKEESKETLIRRLGEPE
DCAGIVSEFCSDEDASYITGEFVVGGPSPRL"
3'UTR
polya_signal
BASE COUNT 957 962
230 a 250 c 322 g 201 t
ORIGIN

alignment_scores:
Quality: 923.00 Length: 278
Ratio: 4.807 Gaps: 1
Percent Similarity: 69.065 Percent Identity: 69.065

alignment_block:
US-09-866-034-2 x AY071856

Align seg 1/1 to: AY071856 from: 1 to: 1003

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTyrAsnSerVa 17
6 ATGCACAAAGCGGGGCTAGGCGCTGTGTGCCGGGCTTGGAATTCGCT 55
17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
56 GCGGATGGCCAGCTCCGGATGACCCGCCGAGCCGCTCGCAAAATAGG 105
34 aLLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
106 TGGCCCTGTAAGCGCTCCACCGAGGATCGGCTTCCCATGGCCCG 155
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
156 CGTTGGCCAGAGACGGGCGCCATGTGCTGTCAGCAGCGGAAGACGA 205
67 AsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
206 GAATGTGACACAGCGGCTGCCAGCTGCAGGGGAGGGGCTCAGCGTGA 255
84 hGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
256 CGGGCACCGTGTCCATGTGGGAAGCGGAGGAGCGGCGTGGTG 305
101 AlaThrAlaValLysLeuHISglyIleAspIleLeuValSerAsnAl 117
306 GCCACGGCTGTGAAGCTTCATGAGGATCGAATCTCTCCATGC 355
117 aLValAsnProPhePheGlySerIleMetAspValThrGluIuValT 134
356 TCGTGCACACCTTCTTGGAAAGCATATGATGTCACAGAGAGGTGT 405
134 rAspThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
406 GGGCAGAG..... 413
151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleVal 167
413 ..... 413
167 aLSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
413 ..... 413
184 aLSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200

```

```

/db_xref="GI:13938402"
/translation="MSSTGIDRKGVLNRYVAVVTGSGIGFALIRPADGHHVVIS
SRKQONDRAMAKLQEGSLVAGIVCHYKGAEDREDDYAKALEHCGVDPLVSGAVN
PLNGSTIGTSEQIWDKTLSTNVKSPALLSÖLLPMEYNRGAIVLSIAIYPPVAL
GVNVSTALDGLRTALELAPDIRVNCVVGIIKTDFSKVFHGNESIMKFEHH
ÖLOMIGSEDCAGIVSFLCSPDASYGVNGENIAVAGYSTRL"
BASE COUNT      271 a      279 c      398 g      278 t
ORIGIN

alignment_scores:
      Quality: 833.50      Length: 272
      Ratio: 3.502      Gaps: 3
Percent Similarity: 87.500      Percent Identity: 61.029

alignment_block:
US-09-866-034-2 x BC007339      ..

Align seg 1/1 to: BC007339 from: 1 to: 1226

2      HSLYSLAGIYLEUENGLYLEU     CYALAIARYALATRPASNSERVALA 18
      |||:|||||:|||||:|||||:|||||
3      CAGGAGGCGAGGCGTGTTCATCCCTGTGCAGAGCTT.....TCTGTGA 46
      |||:|||||:|||||:|||||:|||||

18     rmetAlaSerSerSergIymethrArgrArGAPProLeuAlaAsnLysVal 34
      |||:|||||:|||||:|||||:|||||
47     GGAATAGGACGACCGCGGATAGACAGAAAGGCGCTGTGCTPACCGGGSTA 96
      |||:|||||:|||||:|||||:|||||

35     AlAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaIar 51
      |||:|||||:|||||:|||||:|||||
97     GCCGGGTGCACGGGGGTCCACACAGTGGATCGCTTGGCATCGCCGACG 146
      |||:|||||:|||||:|||||:|||||

51     glenUlaGlnAspGlyAlaHisValValaSerSerArgLysGlnGlna 68
      |||:|||||:|||||:|||||:|||||
147    TCTGGCCGGGGACGGGGGCCACGTGGTCATAGACAGCGGAAAGCAGAGA 196
      |||:|||||:|||||:|||||:|||||

68     snValAspGlnAlaValAlaThrLeuGlnGlyGlnGlyLeuSerValThr 84
      |||:|||||:|||||:|||||:|||||
197    ACGTGGACGGGGCCATGGCCCAAGCTGCAGGGGGAGGGGCTAGTGTGGCG 246
      |||:|||||:|||||:|||||:|||||

85     GlyThrValCysHisValGlyLysAlaGlnAspArgGluArgLeuValAl 101
      |||:|||||:|||||:|||||:|||||
247    GGCATTGTGTGCCACGTGGGGAAGCTGTAGACCGGAGCAGCTGGTGCC 296
      |||:|||||:|||||:|||||:|||||

101    aThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAla 118
      |||:|||||:|||||:|||||:|||||
297    CAAGGCGCTGGAGCACGTGTGGGGCGTGCAGCTTCTGTGTGTGGCCGAC 346
      |||:|||||:|||||:|||||:|||||

118    lAlaValAspProPheGlySerIleMetAspValIrnGlnGlnValTrp 134
      |||:|||||:|||||:|||||:|||||
347    GGGTCAACCCCTCTGTGTAGGAGACACTGTGGGAGCAATGTAGCAATGTGG 396
      |||:|||||:|||||:|||||:|||||

135    AspIyThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAl 151
      |||:|||||:|||||:|||||:|||||
397    GACAAGATCCYAAGTGTGAAGCTGAAGTCCCAAGCCGTGCTGCAGACCA 446
      |||:|||||:|||||:|||||:|||||

151    aValValProLumetGluLysArgGlyGlySerValValIleValS 168
      |||:|||||:|||||:|||||:|||||
447    GTTCTGTCCTTACATGGAGAACAGG...AGGGGTGCTGTATCTCTGTCT 493
      |||:|||||:|||||:|||||:|||||

168    eSerIleAlaAlaPheSerProSerProGlyPheSerProTyArgVal 164
      |||:|||||:|||||:|||||:|||||
494    CTTCATTTGACGCTTATATATCACTAATGTGGCGGTGTCTCAAAATGTC 543
      |||:|||||:|||||:|||||:|||||

185    SerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAl 201
      |||:|||||:|||||:|||||:|||||
544    AGCAAGACAGCGCTGTGGGTCTCAGTAAACAACGTGCATTGAGAGCTGGC 593
      |||:|||||:|||||:|||||:|||||

201    aProArgAsnLleArgValAsnGlyLeuAlaProGlyLeuIleLysThrS 218
      |||:|||||:|||||:|||||:|||||
594    CCCCAAGACATCCGGGTAAACTGCGTGTCCAGCAATATCAAAACTG 643
      |||:|||||:|||||:|||||:|||||

218    eRPheSerArgMetLeuTrpMetAsnLysGlnLysGlnGlySerMetIre 724
      |||:|||||:|||||:|||||:|||||

```



```

644 ACTTCACCAAGTGTTCATGGGAAATGAGTCTCTGTGAAGAACTCAAG 693
235 GltuHrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIle 251
694 GAAACATATAGAGCTGACGAGAGATGGGAGCTGAGAGCTGTGCAGGAAAT 743
251 eValSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrV 268
744 CGGTGCTCTCTGCTGCTCTCCAGATGCGACACTAGCTACAGGGGAGAAC 793
268 aValValAlaGlyGly 272
794 TTGCGGTGCGCAGCG 807

seq_name: gb_pr:HSU31875
seq_documentation_block:
LOCUS HSU31875 1442 bp mRNA linear PRI 29-NOV-1995
DEFINITION Human Hep27 protein mRNA, complete cds.
ACCESSION U31875
VERSION U31875.1 GI:1079565
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1442)
Gabrielli, F., Donadel, G., Bensl, G., Heguy, A. and Mell, M.
A nuclear protein, synthesized in growth-arrested human
hepatoblastoma cells, is a novel member of the short-chain alcohol
dehydrogenase family
Eur. J. Biochem. 232 (2), 473-477 (1995)
96035881
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
Biochemistry, University of Pisa, Via Roma 55, Pisa 56126, Italy
FEATURES
SOURCE
1. 1442
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="hepatoblastoma Hep2"
434. 1276
/note="similar to Streptomyces violaceoruber granatcin
polyketide synthase putative ketocycl. Swiss-Prot
Accession Number P16542"
/product="Hep27 protein"
/codon_start=1
/protein_id="AA82048.1"
/db_xref="GI:1079566"
/translation="MLSAVVARGYGWPHPCARLSVRMSSTGIDRRGVLANVAVVTGS
TSIGIFAIARLARDGAHVIVSSRKQNVDRAMKLOGEGISVAGIYCHVGAEDREG
LVAKALEHCGGVDFVLCAGVNPPLVSGISLIGTSPQIMDKILISVANKSPALLSOLPLPT
ENRRGAVILVSSIAVNPVVALGVYNSKTLALDITRLALELAPKDIRVCVYPGIT
KTDRSKVPHGHESLMKNPKEHHQORIGESDCAGIYSFLCSPASIVYNGENIYVAGI
STR."
BASE COUNT 322 a 369 c 440 g 311 t
ORIGIN
alignment_scores:
Quality: 824.50 Length: 263
Ratio: 3.569 Gaps: 2
Percent Similarity: 87.833 Percent Identity: 61.597
alignment_block:
US-09-866-034-2 x HSU31875
Align seg 1/1 to: HSU31875 from: 1 to: 1442
10 CysAlaArgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrAr 26

```

```

479 TGTGTCTAGAGCTT.....TCTGTAGAGATTGACACACCGGATAGACAG 522
26 gArGAspProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspG 43
523 GAAGCGGCTCTGAGCTAACCGGATAGCGGTGTACAGGGGTCCACAGATG 572
43 LylleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAlaHisVal 59
573 GGAATCGGCTTGGCATGCGCCGAGCTGTGGCCGGGAGCGGCGCCAGCTG 622
60 ValValSerSerArgLysGlnGlnAsnValAspLysAlaValAlaThrLe 76
623 GTCAATCAGACGCGGAACAGACAGATGGACCGGACGAGCGCAAGCT 672
76 uGlnGlyGlnGlyLeuSerValThrGlyThrValLysHisValAlaLysA 93
673 GCAGGGGAGAGGGCTGAGTGTGGCGGCAATGTGTGCGACAGTGGGGAAG 722
93 laGluAspArgGluArgLeuValAlaThrAlaValLysLeuHisGlyGly 109
723 CTGAGGACCGGAGCAGCTGTGTGCCAAGGCCCTGGAGACACTGTGGGGGC 772
110 IleAspIleLeuValSerAsnAlaAlaValAsnProPheNegIleSerI 126
773 GTCACTTCTCTGTGTGCACAGCGAGGGGTCAACCTGTGTAGAGACAC 822
126 eMeAspValThrGlnGlnValAlaTrpAspLysThrLeuAspIleAsnVal 143
823 TCTGGGACACAGTACAGACATCTGGGACAGATCTTAAGTGTACAGCTGA 872
143 ySalAspProAlaLeuMetThrLysAlaValAlaProGluMetGlyLysArg 159
873 AGTCCCGACCGCTGCTGTGACGCCAGCTGCTGCCCTCATGTAGAGAACAG 922
160 GlyGlyLysSerValAlaIleValSerSerIleAlaAlaPheSerProse 176
923 ...AGGAGTCTGATCTGTGTCTCTTCATTTGACACTTAATATCAAGT 969
176 rProGlyPheSerProTyrAsnValSerLysThrAlaLeuLeuGlyLeuT 193
970 AGTGGCGCTGGGTGTCTACAAATGTACAGACAGACAGCGCTCTGTGCTCA 1019
193 hLysThrLeuAlaIleGlnLeuAlaProArgAsnIleArgValAsnCys 209
1020 CTAGAACACTGTGATGAGACTGCGCCCAAGACATCCGCGTAACCTGC 1069
210 LeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeuTrpMetAs 226
1070 GTGATTCAGGAATTAATAAACTGACTTCAGCAAGGTGTTCATGGGAA 1119
226 pLysGluLysGlnGluSerMetLysGluThrLeuArgIleArgArgLeu 243
1120 TGAGTCTCTCTGTGAAGAAGCTCAAGAACATCATCAGCTGCACAGAGATTG 1169
243 LysIleProGluAspCysAlaGlyIleValSerPheLeuCysSerGluAsp 259
1170 GGGAGTCAGAGAGACTGTGCAGGAATCGTGTCTCTCTGAGCTCCAGAT 1219
260 AlaSerTyrIleThrGlyGluThrValValValAlaGlyGly 272
1220 GCCAGCTACGTCAACGGGAGAACATTTGGGTGGCAGCG 1258

seq_name: gb_in:AY069779
seq_documentation_block:
LOCUS AY069779 1113 bp mRNA linear INV 17-DEC-2001
DEFINITION Drosophila melanogaster SD02021 full length cDNA.
ACCESSION AY069779
VERSION AY069779.1 GI:17862893
KEYWORDS FLI cDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```

REFERENCE

Eukaryota: Melazoa: Arthropoda: Tracheata: Hexapoda: Insecta: Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha: Ephydroidea: Drosophilidae: Drosophila. 1 (bases 1 to 1113)

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C. J., Nuncio, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M. and Celniker, S.

TITLE

Submitted (10-DEC-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

COMMENT

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcriptase errors that resulted in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

Location/Qualifiers

1. 1113
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/map="64C12-64C12"
1. 1113
/gene="CG10672"
/note="alignment with genomic scaffold AE003567"
/db_xref="FLYBASE:FBgn0035588"
105. 1058
/gene="CG10672"
/note="Longest ORF"
/codon_start=1
/db_xref="SD02021P"
/product="SD02021P"
/protein_id="AAL39924.1"
/db_xref="GI:17862894.1"
/translation="MLHLSKQLVVRAPKIRLNASVSGSSSPDONSNNYSRKLVG
P/LNQC/KRLSSSSOSTAGTMR/LAGKAVVAVVASTDGI/FA/LAKRLAEGCAVVIS
RKQKND/SALAE/LKLN/LNVH/GKCHVSEDEDK/LEEFIS/GKLNILVSAATNP
AVGVLECDK/VWDKIEDVNVKSS/LYLAKALP/LRQKRSIVP/SSLAGYAFEL
GAYSVKTA/LIGTKRAAND/LAPBGIRV/KNLACGVR/TRKSK/LYENAS/NEALSKT
PMKRLT/SEMAVGSF/VSEDGVTG/ESTIVAGGWTARL"

CDS

BASE COUNT 308 a 266 c 300 g 239 t
ORIGIN

alignment_scores:
Quality: 664.00 Length: 267
Ratio: 3.103 Gaps: 1
Percent Similarity: 80.150 Percent Identity: 50.936

alignment_block:
us-09-866-034-2 x AY069779

Align seg 1/1 to: AY069779 from: 1 to: 1113

12 ATGATAATpAsnSerValArgMetAlaSerSerglyMetThrArgArgAs 28
111
261 CCCTTGTCACGACGATGCTCAAGCTTCGACGCCGACGATGACGCT... 308

28 pProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGlyIleG 45
309 ... TTGCGACGAAAGTAGCCCTGGTCACGCCCTCCACCGATGCATTG 354
45 lYpHeAlaIleAlaArgArgLeuAlaGlnAspGlyValAlaHisValAl 61
355 GTTTCCGATTGCCAAGAGCTGGCCGAGATGGCCCGACGATGGTCATC 404
62 SerSerArgLysGlnGlnAsnValaAspGlnAlaValAlaThrLeuGlnI 78
405 ACCAGTCGCAACGACAAAGATGTGACTCTGCTGCGGAGCTGGCA 454
78 yGlLysLeuSerValThrGlyThrValCysHisValGlyLysAlaGlu 95
455 ACTGAACCTGAATGTCCACGGTCTGAAAGCCATGTCACGAAACCGAG 504
95 sPaArgLysLeuValAlaThrAlaValLysLeuHisGlyIleAsp 111
505 ATCGCAAGAGCTCTCGAGAGACCATCAGCAAGTTTGCAAGCTTAAC 554
112 lLeuValSerAsnAlaValaAsnProPheGlySerIleMetAs 128
555 ATTCTGTCACGATGCACCCACCAATCTCGCGTGGCGGTGTCGCA 604
128 pValThrGlnGluValTrpAspLysThrLeuAspIleAsnValLysAla 145
605 GTCCGACGAGAGGTGTGGCAAGATCTTCATCGAAGCGAAGAGTT 654
145 rAlaLeuMetThrLysAlaValaProGluMetGluLysArgGlyIle 161
655 CCTATGCTGGCCAAAGAGGACCTGCTCTCTCGCCGACGAAAGAAC 704
162 GlySerValValIleValSerSerIleAlaAlaPheSerProSerProI 178
705 TCCACGATGCTTTCTCCTCCTCCATGCTGGCTATGATGCTTTGAGCT 754
178 yPheSerProTyrAsnValSerLysThrAlaLeuLeuGlyLeuThrLysT 195
755 ACTGGAGGACCTATTCCTCAGCAAGACCGCGCTGATGGCTTGACCAAG 804
195 hLeuAlaIleGluLeuAlaProArgAsnIleArgValaAsnLysLeuAla 211
805 CAGCCGCAAGGATCTGGCCCGGAGGCGCATTCGCTCAACTGCTGGCT 854
212 ProGlyLeuIleLysThrSerPheSerArgMetLeuTrpMetAspLysG 228
855 CCAGAGATCATCAGCACAAGTTCTCGAAGCAGCTGACAGAAATGATGC 904
228 uLysGlnGluSerMetLysGluThrLeuArgIleArgArgLeuGluP 245
905 GCGCATGTAAGCGGCTTGACCAAAATACCATTGGCTGCTGGCAGCA 954
245 rGluAspCysAlaGlyIleValSerPheLeuCysSergLysAlaSer 261
955 GCGAGAGATGGCTGGCGGCTCTCTTTGGTCCGAGATGGGAG 1004
262 TyrIleThrGlyLysThrValValaGlyGlyThrProSerArg 278
1005 TACATTACCGGAGAGTCATTGTGGCGGCGGTGGAATGACAGCTCGTTT 1054
278 u 278
1055 G 1055
seq_name: gb_htg:AC015075
seq_documentation_block:
LOCUS AC015075 34084 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
pieces.
ACCESSION AC015075
VERSION AC015075.1 GI:6436260
KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 34084)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT This sequence was identified as CDM:1021081 by the submitter. For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1..34084
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 9185 a 7818 c 7622 g 9459 t
ORIGIN
alignment_scores:
Quality: 619.00 Length: 309
Ratio: 2.879 Gaps: 3
Percent Similarity: 69.579 Percent Identity: 44.984
alignment_block:
US-09-866-034-2 x AC015075/rev ..
Align seg 1/1 to reverse of: AC015075 from: 1 to: 34084
12 ArgAlaTTPaNSerValArgMetAlaSerSerglyMetThrArgArgas 28
17151 CCTTGTCTCAGAGTAAAGTTGACAGCGCGCAGCATGAAGCGT... 17104
28 proLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGly... 43
17103 ...TTGGCAGGAAAGTACCGGTGTACCGCTCCACCGATGGGTAAAG 17058
43 43
17057 CAGCTCAATAGAACTGAGAGATGCCACCACTAATTCGATTATATC 17008
44 IleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAl 57
17007 CCTCCAGCATGTGGTTCCGATTGCCAAAGCGTCGCCAGATGGCGC 16958
57 aHisValValAlaSerSerArgLysGlnGlnAsnValAspGlnAlaValA 74
16957 CCGAGTGGTATCAGAGTCGCAAGCAAAAGATGTGGACTGTGCTGTGG 16908
74 laThrLeuGlnGlyGlnGlyLeuSerValThrGlyThrValCysHisVal 90
16907 CGGAGCTGGCGCAACGAACTGAAATGTCCAGGCTGTGAAGTGCATGTC 16858
91 GlyLysAlaGlnAspArgGlnArgLeuValAlaThrAlaValLysLeuH 107
16857 ACGGAAACCGAGATCGCAACAGCTCTTGAAGAGACCATTAAGCAAGTT 16808
107 sGlyGlyIleAspIleLeuValSerAsnAlaAlaValAsnProPhePheG 124
16807 TGGCAAGCTCAACATTTGGTTCAGCAATGCACCAATCCTGCGGTGC 16758
124 lYSerIleMetAspValThrGlnGluValTTPAspLysThrLeuAspIle 140
16757 GCGGTGTCTCAGTCGATGCGAAGAGTGTGGACACAGATCTTCGATGTG 16708
141 AsnValLysAlaProAlaLeuMetThrLysAlaValValProGluMetG 157
16707 AACGTGAAGAGTCTCTATCTGTGCGCAAGAGGACGACCTCTCTCTGCG 16658

157 uLysArgGlyGlyLysSerValIleValSerSerIleAlaAlaPheS 174
16657 CCAGCAAAAGAACTCCAGCATGCTTTCGTCTCTCATTTGCTGGCTATG 16608
174 erProSerProGlyPheSerPro..... 181
16607 ATGCTTTTGA.GGTTACTGCTTATCTGCGGTTCAGTTGTCAATGCTT 16559
182 TyrAsnValSerly 186
16558 ATCTCCGTATTCTTATTCGATACCTACTGAGACCTATTTCGACGAA 16509
186 sThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaPro 203
16508 GACCGCGCTGATTGGCTTGACCAAGCAGCGCCGCAAGATCTGGCGCGG 16459
203 rGAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSerPhe 219
16458 AGGCGATTCGCTCACTGCTGCTCCAGAGATCATCAGACAAAGTTTC 16409
220 SerArgMetLeuThrPheMetAspLysGluLysGlnGlnSerMetLysGlu 236
16408 TCGAAACCACTGTACGAGATGAGTCGCGCAATGAGCGGCTTTGACCAA 16359
236 rLeuArgIleArgArgLeuGlyLysProGluAspCysAlaGlyIleValS 253
16358 AATACCATGTGCTGCTGCGCACAGTGAAGATGCGTGGCGGTGCT 16309
253 erPheLeuCysSerGluAspAlaSerTyrIleThrGlyLysThrValVal 269
16308 CTTTGTGCTCTCGAGGATGCGGCTTACATTACCGAGATCATTTG 16259
270 ValGlyGlyGlyThrProSerArgLeu 278
16258 GCTGGGCTGATGATGACAGCTCTTTG 16232

OM of: US-09-866-034-2 to: N_Geneseq_032802.* out_format : pfs

Date: May 9, 2002 3:13 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09866034/rnmat.08052002.081344.16474/app-query.fasta.1.338
-DB=N_Geneseq_032802 -QPMF=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-QCAPOP=4.500 -QCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsun62
-TRANS-human0.cdi -LISTN=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09866034.ecgn1.1.573 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-866-034-2

Query length: 278

Database: N_Geneseq_032802.*

Database sequences: 1736436

Database length: 858457221

Search time (sec): 256.080000

score_list:

Sequence	Strid Orig	ZScore	EScore len	Documentation
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA51259 +		1399.00	2488.49	2.9e-130
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA42131 +		1399.00	2488.34	3.0e-130
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH1816 +		1395.00	2481.53	7.1e-130
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA426090 +		1327.00	2359.42	4.5e-123
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH07482 +		976.00	1736.60	2.2e-88
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH06587 +		906.00	1604.47	5.1e-81
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA540887 +		906.00	1604.47	5.1e-81
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA529730 +		906.00	1604.47	5.1e-81
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAE21725 +		722.50	1281.16	5.2e-68
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL16757 +		664.00	1172.98	5.5e-57
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL16756 +		619.00	1080.70	7.6e-52
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL07994 +		617.00	1078.70	9.8e-52
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC46747 +		588.50	1036.64	2.2e-49
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH3985 +		483.50	883.84	3.3e-39
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA529623 +		434.00	763.81	3.4e-34
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA529624 +		434.00	761.29	4.7e-34
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA506586 +		408.00	722.34	6.9e-32
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA540911 +		408.00	722.34	6.9e-32
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA529729 +		408.00	722.34	6.9e-32
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA245749 +		404.00	710.04	3.3e-31
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA536357 +		377.50	661.92	1.6e-28
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH54351 +		375.00	641.68	2.1e-27
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH8431 +		371.00	587.44	2.3e-24
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC02120 +		370.00	658.87	2.4e-28
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA552507 +		368.50	644.96	1.4e-27
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAH61016 +		365.50	602.92	3.1e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAH05110 +		365.50	602.92	3.1e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV30458 +		362.50	567.83	2.8e-23
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV30459 +		362.50	567.80	2.8e-23
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH81456 +		360.00	588.51	2.0e-24
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH21612 +		360.00	567.72	2.8e-23
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH2554 +		355.50	562.35	5.6e-23
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH63301 +		353.00	618.03	4.5e-26
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH67301 +		349.00	609.78	1.3e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH71517 +		349.00	608.33	1.5e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH71518 +		349.00	608.33	1.5e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH65531 +		349.00	567.99	3.6e-22
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH40009 +		348.50	606.75	1.9e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH51824 +		348.00	609.75	1.3e-25
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH74685 +		348.00	609.71	1.3e-25

seq_name: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA51259

seq_documentation_block:

ID	AA51259 standard; cDNA: 1283 BP.
XX	AA51259;
AC	26-SEP-2000 (first entry)
DT	Human DNA35672-2508 encoding PRO1800, a Hep27 homologue.
XX	PRO1800; Hep27; homologue; short-chain alcohol dehydrogenase; SCAD;
XX	secreted protein; transmembrane protein; recombinant production;
KW	gene therapy; ss.
KW	Homo sapiens.
OS	Location/Qualifiers
XX	36..872
XX	/*tag= a
FT	/product= PRO1800
FT	/note= "The protein is a putative Hep27 homologue"
FT	36..80
FT	/*tag= b
XX	sig_peptide
XX	WO200036102-A2.
XX	22-JUN-2000.
PD	01-DEC-1999; 99WO-US28634
PF	16-DEC-1998; 98US-0112851
PR	16-DEC-1998; 98US-0113145
PR	22-DEC-1998; 98US-0113511
PR	12-JAN-1999; 99US-0115558
PR	12-JAN-1999; 99US-0115733
PR	09-FEB-1999; 99US-0119341
PR	10-FEB-1999; 99US-0119337
PR	12-FEB-1999; 99US-0119665
PR	02-JUN-1999; 99WO-US12252
XX	(GETH) GENENTECH INC.
XX	Boistein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI	Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI	Wood WI;
XX	WPI: 2000-431586/37.
DR	P-PSDB; AAY96729.
XX	Isolated nucleic acid molecule encodes a PRO polypeptide which is a
PT	transmembrane polypeptide
PT	transmembrane polypeptide
XX	Claim 1; Fig 1; 154pp; English.
CC	This CDNA encodes PRO1800, a human Hep27 protein homologue, which has
CC	recently been shown to have homology to known short-chain alcohol
CC	dehydrogenase (SCAD) family of proteins. The invention concerns novel
CC	secreted and transmembrane proteins, designated PRO polypeptides. The
CC	cDNA and gene sequences are useful in the recombinant production of PRO
CC	polypeptides, as a hybridization probe to screen libraries to isolate
CC	cDNAs with sequence identity to PRO polypeptides or to map the gene
CC	encoding the PRO polypeptides and analyzing genetic disorders. The
CC	cDNA/gene can also be used to produce transgenic animals useful for the
CC	development and screening of therapeutically useful reagents. They can
CC	also be used in gene therapy, e.g. to replace a defective gene.

Align seg 1/1 to: AA242131 from: 1 to: 1302

```

1 MetHisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVal 17
34 ATGCACAAAGCGCGCTCTAGCCCTCTGCGCGGCTTGAATTCGGCT 83
17 LArgMetAlaSerSerglyMetThrArgArgAspProLeuAlaAsnLysV 34
84 GCGGATGGCCAGCTCCGGATGACCCCGGAGACCGCTGCAAAATTAAG 133
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
134 TGGCCCTGGTAAAGCGCTCCACCGAAGGATGCGCTTCGCCATTCGCCGG 183
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
184 CTTTGGCCAGGAGCGGCGCATGCTGCTGACAGCAGCGGAGACAGCA 233
67 nasnValAspGlnAlaValAlaThrLeuGlnIleGlyLeuSerValT 84
234 GAATGTGGACCAAGCGGCTGCGCACGCTGCAAGGAGGCGGCTGACGTA 283
84 hGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
284 CGGACCGCTGTCATGTCGGGGAAGCGGAGACCGGAGCGGCTGCTG 333
101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
334 GCGACGGCTGTAGAGCTTCATGAGATGATGATTCATGCTCCAAATGC 383
117 aAlaValAsnProPheGlySerIleMetSerAspAlaThrGluValT 134
384 TGCTGTCAACCTTCTTGGAAAGCATATGATGATGATGAGGAGTGT 433
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
434 GGGACAAAGCTCTGGACATTATGTAAGGCCCGACCCCTGATGACAAAG 483
151 AlaValValProGluMetGluLysArgGlyGlySerValIleLeuVal 167
484 GCAGGAGTCCAGAAATGAGAAAGAGAGAGCGGCTCAGTGTATGATC 533
167 LserSerIleAlaIlePheSerProSerProGlyPheSerProTyrAsnV 184
534 GCTTTCATAGCAGCTTCATGCTCCTGCTGCTGCTGCTGCTGCTGCTG 583
184 aLserLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
584 TCAGTAATAACAGCTTGGCTGGGCTGACCAACACCTGGCCATAGAGCTG 633
201 AlaProArgAsnIleArgValAsnGlyLeuAlaProGlyLeuIleLysTh 217
634 GCCCAAGAGACATTAGGCTGAACTGCCAGCACTGAGCTTATCAAGAC 683
217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysMetL 234
684 TAGCTTCACAGAGATGCTTGGATGACACAGAAAGAGGAGAAACATTA 733
234 yGluThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGly 250
734 AAGAAACCTCGGATTAAGAGATTAGGAGACCAAGAGATTCCTGCTGCG 783
251 LLeuAlaSerPheLeuGlySerGluAspAlaSerTyrIleThrGlyGluTh 267
784 ATCGGTCTTCTTCCTGCTGCTGCAAGATGCCACATACATCGGGAAC 833
267 rValValValGlyGlyGlyThrProSerArgLeu 278
834 AGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 867

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH14816
seq_documentation_block:

ID AAH14816 standard; CDNA, 1258 BP.
XX
AC AAH14816;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12619.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036;
PR 27-AUG-1999; 99JP-0300253;
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WP: 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12619; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC complementary strand of a polynucleotide comprising a sequence complementary to the
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SO Sequence 1258 BP; 282 A; 322 C; 394 G; 260 T; 0 other;

alignment_scores:

Quality: 1395.00 Length: 278
Ratio: 5.018 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.640

alignment_block:

US-09-866-034-2 x AAH14816

Align seg 1/1 to: AAH14816 from: 1 to: 1258

XX	AAZ46090:	
XX		
DT	05-MAY-2000 (first entry)	
XX		
DE	Short chain alcohol dehydrogenase-related molecule ScRM-1 cDNA.	
XX		
KM	Human: short chain alcohol dehydrogenase-related molecule;	
KM	SCAD-related molecule: ScRM-1; ScRM-2, metabolic regulator;	
KM	cell proliferation regulator; inflammation regulator;	
KM	cell proliferative disorder; immune disorder; arteriosclerosis;	
KM	atherosclerosis; bursitis; cirrhosis; hepatitis; AIDS;	
KM	Addison's disease; adult respiratory distress syndrome; allergy;	
XX	ankylosing spondylitis; amyloidosis; cancer; ss.	
OS		
XX	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	CDS	31..867
FT		/*tag= a
FT	/*product=	"short chain alcohol dehydrogenase-related
FT		molecule ScRM-1"
PN	MO200004135-A2.	
PD	27-JAN-2000.	
XX		
PF	16-JUL-1999:	99MO-US16164.
XX		
PR	16-JUL-1998:	9805-0116750.
PR	16-JUL-1998:	9805-0160074.
XX		
PA	(INCYTE) INCYTE PHARM INC.	
PI	Bandman O, Tang YT, Corley NC, Azimzai Y, Baughn MR;	
XX	WPI; 2000-171266/15.	
DR	P-PSDB; AAY68735.	
XX		
PT	New short chain alcohol dehydrogenase polypeptides useful for	
PT	diagnosis, treatment and prevention of cell proliferative disorders	
XX	such as atherosclerosis, cirrhosis and cancers of various tissues	
XX	Claim 7, Fig 1A-D: 78pp; English.	
XX		
CC	The present sequence encodes a human short chain alcohol dehydrogenase	
CC	(SCAD)-related molecule designated ScRM-1. The specification also	
CC	describes ScRM-2. ScRM proteins are metabolic, cell proliferation and	
CC	inflammation regulators. The ScRM polynucleotides and polypeptides are	
CC	used for treating or preventing a cell proliferative or immune disorder	
CC	in humans. Cell proliferative disorders include arteriosclerosis,	
CC	atherosclerosis, bursitis, cirrhosis, and hepatitis. Immune disorders	
CC	include AIDS, Addison's disease, adult respiratory distress syndrome,	
CC	allergies, ankylosing spondylitis, and amyloidosis. The vectors,	
CC	agonists, antagonists, antibodies and complementary sequences are also	
CC	used for treating the above conditions. The polynucleotides and	
CC	polypeptides are also used for treating cancers of various tissues	
XX	such as adrenal gland, bladder, bone, bone marrow, and brain.	
XX		
SQ	Sequence 1280 BP; 301 A; 326 C; 393 G; 260 T; 0 other;	
	Alignment_scores:	
	<Quality: 1327.00	Length: 278
	Ratio: 4.879	Gaps: 0
	Percent Similarity: 97.842	Percent Identity: 96.403
	Alignment block:	
	US-09-866-034-2 x AAZ46090	
	Align seg 1/1 to: AAZ46090 from: 1 to: 1280	

```

||||| ||| ||||||||||||||| ||| ::|||
31 ATCCACATGGCCAGGCTGCTAGGCTCTGTGCTGGGACGAGTGGT 80
17 LARGMETALASERSEGLYMETHRARGARGASPROLEUALASNLVY 34
81 GCGGATGGCCAGCTCAGATGACCCGCGGACCCGCTCACAAATRAG 130
34 ALALEUVALTHRALASERTRASPGLYLEGLYPHEALALEIARG 50
131 TGGCCCTGGTAACGGCTCCACCGACGGGATCGGCTTCGCATCGCCG 180
51 ARGLEUALGLASPGLYALAHISVALVALASERSEARGYSGLNGI 67
181 CGTTTGCCGACGACGAGCCGACCGTGTGCTGACGACCGGAGGACGA 230
67 NASNVALASPGIALAVALATHREUGLNGIYGLUGLYLEUSEVALT 84
231 GAATGTGACACAGCGGCTGCCGCTGCAGGGGAGGGGCTAGCGTGA 280
84 HNGLYTHRALCYSHISVALGLYVALAGLUSPARGLUARGLEUVAL 100
281 CCGGACACCTGTGCCATGTGGGAGGCGGAGACCGGCGGCTGTG 330
101 ALATHRALVALYLSRTHLSGLYGLYLEASPLILEUVALSERASNA 117
331 GCCACGCGCTGTGAGCTTCATGAGGTATCGATCTCTCAATGC 380
117 ALAVALASAPROPHEPHEGLYSERILEMETASVALTHRGLUVALT 134
381 TCGTGTCAACCCCTTCTTGAAGCATATGATGTCACCTGAGAGGTGT 430
134 RPASPLYSTRLEUNSPILLEASVALYALAPROALALEUMETHTLYS 150
431 GGGACAGACTCTGGACATTATGTGAAGGCCGACCTGATGACAAAG 480
151 ALAVALPROGLMETGLULYSARGGLYGLYSERVALLEVALLE 167
481 GCAGTGTGTCAGAAATGGAGAAACAGAGAGGGGCTCAGTGTGATCG 530
167 LERSEIRLEALALAPHERPROSERPROGLYPHERPROTYRASN 184
531 GCTTTCATAGACGCTTCAGTCATCTCTGGCTCAGCTCTTACAATG 580
184 ALSERLYSTRALALEULEUGLYLEUTHLYSTRLEUALALEGLU 200
581 TCAGTAAACACGCTTGTGGGCTTCACAAATACCTGGCCATAGAGCTG 630
201 ALAPROAIRASNILEARGVALASNCYSLEUALAPROGLYLEU 217
631 GCCCAAGGACATTAAGGTGAACCTCTAGACACCTGGACTTATCAAGAC 680
217 RSEPHESERARGMETLEUTRPMETASPLYSGULYSGULUSERMET 234
681 TAGCTTCACACAGATGCTCTGATGACAGAGAAAGAGAAAGCATGA 730
234 YSGULUTHREUARGLIARGARGLEUGLYGLUPROGLUASPCYSALA 250
731 AAGAAACCCCTGGGATTAAGAAAGTTAGCGACGACGAGATGTGCTGC 780
251 ILEVALSERPHELEUCYSERGLUSPALASERTYRILETHGLYLU 267
781 ATCGTGTCTTCTTCTGTGCTGTAAGATGCCACTACATCAGTGGGAAC 830
267 RVALVALAIGLYGLYGLYTHRPROSERARGLEU 278
831 AGTGGTGGTGGTGGAGGAGACCCGCTCCGCTC 864

```

seq_name: /SIDS1/gcgdata/geneseq-emb1/NA2001A.DAT:AAH07482

seq_documentation_block:

ID AAH07482 standard; CDNA: 680 BP.
XX
AC AAH07482;

```

XX 26-JUN-2001 (first entry)
DT Human CDNA clone (5'-primer) SEQ ID NO:4317.
XX
DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
KW Homo sapiens.
XX
OS EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1: SEQ ID 4317; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH35893 represent human amino acid sequences; AAH92446 to
CC AAH13632 represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 680 BP; 148 A; 180 C; 216 G; 130 T; 6 other:

```

alignment_scores:

Quality: 976.00 Length: 218
Ratio: 4.670 Gaps: 2
Percent Similarity: 95.872 Percent Identity: 94.954

alignment_block:

US-09-866-034-2 x AAH07482

Align seg 1/1 to: AAH07482 from: 1 to: 680

1 MethISLALAGLYLEUGLYLEUCYSALAARGALATRPASNSERVA 17
|||||

21 ATGCACAAAGCGGGCTGTAGGCTCTGTGCCCGGCTTGGAAATCGCT 70
17 IArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
71 GCGGATGGCCACGCTCCGGGATGACCCGCGGACCCCTCGCAATATGAG 120
34 AAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
121 TGGCCCTGGCAAGCGCTCCACGACGGGATGGGCTTCGCATCGCCCGG 170
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
171 CGTTTGGCCCGGAGCGGGCCCATGTGTGTCAGACGCGGAGGACGCA 220
67 nasuValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
221 GAATGTGGACAGCGCGGTGGCCACGTCGACAGGGGAGGGGCTGAGCGTGA 270
84 hrGlyThrValCysHisValGlyLysAlaGlnAspArgGluArgLeuVal 100
271 CGGACACCGTGTGCCATGTGGGGAAGCGGAGGACCGGAGCGGCTGTG 320
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAl 117
321 GCCACGGCTGTGAAGCTTCATGGAGGTATGATCTCCTAGTCTCAATGC 370
117 aAlaValAspProPhePheGlySerIleMetAspValThrGluGluValT 134
371 TGGCTGAACCCCTTTCTTGGAGCATATATGATGATGATCTGAGAGGTGT 420
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
421 GGCACAGACTCTGACATTAATGTGAANCCCCAGCCCTGATACAAAG 470
151 AlaValAlaProGluMetGluLysArgGlyGlyLysValValIleVal 167
471 GCACTGTGCAGAAATGAGAAACGAAAGCGGCTCAGTGGATGCT 520
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsn 183
521 GTCTTCATTAAGAAAGCTTCAGTCCATCTCTGGCTTCAGTCTTACAAAT 570
184 ValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLe 200
571 GTCAGTAAACAGCCTTGTCTGGCTGACCAANACCTGCGCATANAGCT 619
200 uAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLys 216
620 GGGCCCAAGGACATTATGCTGAACCTGCTAACACCTGACATWATCAAA 669
seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.ABA06587
seq_documentation_block:
ID ABA06587 standard; cDNA: 1292 BP.
XX
AC ABA06587;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human CDNA SEQ ID NO: 253.
XX
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.

XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUL-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 14-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218250P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226686P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234979P.
PR 25-SEP-2000; 2000US-234988P.
PR 26-SEP-2000; 2000US-234989P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.

PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237033P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239335P.
 PR 13-OCT-2000; 2000US-239337P.
 PR 20-OCT-2000; 2000US-240966P.
 PR 20-OCT-2000; 2000US-241221P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241786P.
 PR 20-OCT-2000; 2000US-241787P.
 PR 20-OCT-2000; 2000US-241808P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 20-OCT-2000; 2000US-241826P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 08-NOV-2000; 2000US-246474P.
 PR 08-NOV-2000; 2000US-246475P.
 PR 08-NOV-2000; 2000US-246476P.
 PR 08-NOV-2000; 2000US-246477P.
 PR 08-NOV-2000; 2000US-246478P.
 PR 08-NOV-2000; 2000US-246523P.
 PR 08-NOV-2000; 2000US-246524P.
 PR 08-NOV-2000; 2000US-246525P.
 PR 08-NOV-2000; 2000US-246526P.
 PR 08-NOV-2000; 2000US-246527P.
 PR 08-NOV-2000; 2000US-246528P.
 PR 08-NOV-2000; 2000US-246532P.
 PR 08-NOV-2000; 2000US-246609P.
 PR 08-NOV-2000; 2000US-246610P.
 PR 08-NOV-2000; 2000US-246611P.
 PR 08-NOV-2000; 2000US-246613P.
 PR 17-NOV-2000; 2000US-249207P.
 PR 17-NOV-2000; 2000US-249208P.
 PR 17-NOV-2000; 2000US-249209P.
 PR 17-NOV-2000; 2000US-249210P.
 PR 17-NOV-2000; 2000US-249211P.
 PR 17-NOV-2000; 2000US-249212P.
 PR 17-NOV-2000; 2000US-249213P.
 PR 17-NOV-2000; 2000US-249214P.
 PR 17-NOV-2000; 2000US-249215P.
 PR 17-NOV-2000; 2000US-249216P.
 PR 17-NOV-2000; 2000US-249217P.
 PR 17-NOV-2000; 2000US-249218P.
 PR 17-NOV-2000; 2000US-249244P.
 PR 17-NOV-2000; 2000US-249245P.
 PR 17-NOV-2000; 2000US-249264P.
 PR 17-NOV-2000; 2000US-249265P.
 PR 17-NOV-2000; 2000US-249297P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 17-NOV-2000; 2000US-249300P.
 PR 01-DEC-2000; 2000US-250160P.
 PR 01-DEC-2000; 2000US-250391P.
 PR 05-DEC-2000; 2000US-251030P.
 PR 05-DEC-2000; 2000US-251988P.
 PR 05-DEC-2000; 2000US-256719P.
 PR 06-DEC-2000; 2000US-251479P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251858P.
 PR 08-DEC-2000; 2000US-251863P.
 PR 08-DEC-2000; 2000US-251989P.
 PR 08-DEC-2000; 2000US-251990P.
 PR 11-DEC-2000; 2000US-254097P.
 PR 05-JAN-2001; 2001US-259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-476161/51.
 DR P-PSDB; ABB10365.
 XX
 XX Isolated nucleic acid molecule encoding an inflammation-associated

PT polypeptide is used in preventing, treating or ameliorating a medical
 PT condition
 XX
 PS Claim 1; SEQ ID NO: 253; 859pp + Sequence Listing; English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a cDNA of the invention.
 XX
 SO Sequence 1292 BP; 282 A; 339 C; 423 G; 238 T; 10 other;

alignment_scores:
 Quality: 906.00 Length: 278
 Ratio: 4.768 Gaps: 1
 Percent Similarity: 68.345 Percent Identity: 67.986

alignment_block:
 US-09-866-034-2 x ABA06587

Align seg 1/1 to: ABA06587 from: 1 to: 1292

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaAlaArgAlaTrpAsnSerVa 17
 263 ATGCACACAGGCSRGCTCTAGGCTCTGCGCCGGCTTGCAATTCGGT 312
 17 largMetIaSerSerGlyMethThrArgArgAspProLeuAlaAsnLysV 34
 313 GCGGATGCCAGCTCCGGATGACCCCGCGGACCCGCTGCAAAATAGG 362
 34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 363 TGCCCTGTGTACGGCTCCACGACGCGATGCGCTTCCATTCGCCCGG 412
 51 ArgLeuAlaGlnAspGlyAlaHisValValIleSerSerArgLysGlnG 67
 413 CATTGGCCCGACGACGCGGCCCACTGTGCTGCACGACCGGAAGCAGCA 462
 67 nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValI 84
 463 GAATGTGACACAGCGCGGCGCAGCTGCAGGGGGGGGGCTGACGCTGA 512
 84 hrcGlyThrValCysHisValGlyIleValAlaGlnAspArgGluArgLeuVal 100
 513 CGGGACCGCTGTCCATGTGGGAAGCGGAGACCGGAGCGCTGCTG 562
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
 563 GCCAYGCGTGTGAAGCTCATGAGGTATCGATCTGATCTCCAAATG 612
 117 aAlaValAsnProPhePheGlySerIleMetAspValThrGluValI 134
 613 TGCTGTCAACCTTCTTGTGAAGCMTAATGATGATGACGAGAGAGCTG 662
 134 rPAspLysThrLeuAspIleAsnValAlaProAlaLeuMetThrLys 150
 663 GGGACACAG..... 670
 151 AlaValAlaProGluMetGluLysArgGlyGlySerValAlIleVa 167
 670 670
 167 ILeSerIleAlaAlaPheSerProSerProGlyPheSerProTyAsnV 184
 670 670
 184 aLserLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
 670 670
 201 AlaProArgAsnIleArgValAsnGlyLeuAlaProGlyLeuIleLysTh 217

```
670 ..... 670
217 rSerPheSerArgMetLeuTrpMetAspIysGluGlnIserNectL 234
671 .....CTGTGGATGGACAAAGAAAAGAGAAAGCATGA 704
234 ysgIuTrIleuArgIleArgArgIleuGlyIuPProGluAspCysAlGly 250
705 AAGAACCCTCGGATAGAGAGTTAGCGGAGCCAGAGATTGTCTGCG 754
251 lIleValSerPheIeuCysSerGluAspAlaSerTyrIleThrGlyIuTh 267
755 ATCGTGTCTTCTGCTGCTGTGAGATGCCAGCATCACTGGGAAAC 804
267 rValValValGlyGlyGlyThrProSerArgIleu 278
805 AGTGTGTGTGTGTGAGAAACCCGCTCCGCTC 838

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA540887
seq_documentation_block:
ID AA540887 standard; cDNA: 1292 BP.
XX
AC AA540887;
XX
DT 17-DEC-2001 (first entry)
XX
DE CDNA encoding novel human enzyme polypeptide #103.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
PN W020015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220966.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
```

```
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
```

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249255.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Rosen CA, Barash SC, Ruben SM:
 XX WPI: 2001-465566/50.
 DR P-PSDB: AAU23017.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 4: SEQ ID NO 113; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 1292 BP: 282 A; 339 C; 423 G; 238 T; 10 other;

alignment_scores:

Quality: 906.00 Length: 278
 Ratio: 4.768 Gaps: 1
 Percent Similarity: 68.345 Percent Identity: 67.986

alignment block:

US-09-866-034-2 x AAS40887

Align seg 1/1 to: AAS40887 from: 1 to: 1292

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
 |||||
 263 ATGCACAGGCSRGCTGCTAGGCTCTGTGCCGGGCTTGGAATTCGT 312
 17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysVa 34
 |||||
 313 GCGGATGCGCAGCTCCGGGATGACCCCGCGGACCCGCTGCCAATRAGG 362
 34 aLaLeuValThrAlaSerThrAspGlyTleGlyPheAlaIleAlaArg 50
 |||||
 363 TGGCCCTGCTAACGGCTCCACGACGGGATCGGCTTCGCCATCCGCCGG 412
 51 ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgLysGlnG 67
 |||||
 413 CGTTGGCCCGAGGAGCGGCCGCAVGTGTCTGCAGCAGCCGGAAGCAGCA 462
 67 nAsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValTr 84
 |||||
 463 GATGTGACACACCGCGTGGCCACGCTGCAGGGGAGGGCTGAGGTGA 512
 84 hArgLysThrValCysHisValGlyLysAlaGlnAspArgGlnArgLeuVal 100
 |||||
 513 CGGGACCGCGTGTCCATGTGGAGAGCGGAGCGGAGCGGCGGTGCTG 562
 101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnL 117
 |||||
 563 GCCATGCTGTGAAGCTTCATGGAGGTATCGATATCTGTATCCCAATGC 612
 117 aAlaValAsnProPhePheGlySerIleMetAspValThrGlnGlnValTr 134
 |||||
 613 TGCTGTCAACCCCTTCTTGGAGCGMTATGATGTGCACGAGAGAGTGT 662
 134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 |||||
 663 GGGACCAAG..... 670
 151 AlaValValProGlnMetGlyLysArgGlyGlySerValValIleVal 167
 670
 167 LserSerIleAlaAlaPheSerProSerProGlyPheSerProTrpAsnV 184
 670
 184 aSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlnLeu 200
 670
 201 AlaProArgAsnIleArgValAsnLysCysLeuAlaProGlyLeuIleLysTh 217
 670
 217 rSerPheSerArgMetLeuTrpMetAspLysGlyGlnGlnSerMetL 234
 |||||
 671CTTGGATGACACAGGAAAGAGAAAGCATATA 704
 234 yGlnLThrLeuArgIleArgArgLeuGlyGlnProGlnAspCysAlaGly 250
 |||||
 705 AAGAAACCTCGGATGATAGAGGTTAGCGGACGACCAAGATTTGTCTGCC 754
 251 lIleValSerPheLeuCysSerGlnAspAlaSerTrpIleThrGlyGlyL 267
 |||||
 755 ATCGTGTCTTCTTCGTGTGATGACATGCACATCATCATGCGGAAC 804

PR	01-SEP-2000	2000US-0229343.4
PR	01-SEP-2000	2000US-0229343.44
PR	01-SEP-2000	2000US-0229343.444
PR	05-SEP-2000	2000US-0229509.00
PR	05-SEP-2000	2000US-0229513.00
PR	06-SEP-2000	2000US-0230437.00
PR	06-SEP-2000	2000US-0230438.00
PR	08-SEP-2000	2000US-0231242.00
PR	08-SEP-2000	2000US-0231243.00
PR	08-SEP-2000	2000US-0231244.00
PR	08-SEP-2000	2000US-0231414.00
PR	08-SEP-2000	2000US-0232086.00
PR	08-SEP-2000	2000US-0232081.00
PR	12-SEP-2000	2000US-0231968.00
PR	14-SEP-2000	2000US-0232337.00
PR	14-SEP-2000	2000US-0232338.00
PR	14-SEP-2000	2000US-0232339.00
PR	14-SEP-2000	2000US-0232400.00
PR	14-SEP-2000	2000US-0232401.00
PR	14-SEP-2000	2000US-0233063.00
PR	14-SEP-2000	2000US-0233064.00
PR	14-SEP-2000	2000US-0233484.00
PR	14-SEP-2000	2000US-0233485.00
PR	21-SEP-2000	2000US-0234272.00
PR	21-SEP-2000	2000US-0234274.00
PR	25-SEP-2000	2000US-0234997.00
PR	25-SEP-2000	2000US-0234998.00
PR	25-SEP-2000	2000US-0234999.00
PR	26-SEP-2000	2000US-0235484.00
PR	27-SEP-2000	2000US-0235834.00
PR	27-SEP-2000	2000US-0235835.00
PR	29-SEP-2000	2000US-0236327.00
PR	29-SEP-2000	2000US-0236327.00
PR	29-SEP-2000	2000US-0236357.00
PR	29-SEP-2000	2000US-0236358.00
PR	29-SEP-2000	2000US-0236359.00
PR	02-OCT-2000	2000US-0236807.00
PR	02-OCT-2000	2000US-0236808.00
PR	02-OCT-2000	2000US-0237037.00
PR	02-OCT-2000	2000US-0237039.00
PR	02-OCT-2000	2000US-0237303.00
PR	02-OCT-2000	2000US-0237304.00
PR	13-OCT-2000	2000US-0239395.00
PR	13-OCT-2000	2000US-0239397.00
PR	20-OCT-2000	2000US-0240926.00
PR	20-OCT-2000	2000US-0241191.00
PR	20-OCT-2000	2000US-0241192.00
PR	20-OCT-2000	2000US-0241785.00
PR	20-OCT-2000	2000US-0241786.00
PR	20-OCT-2000	2000US-0241787.00
PR	20-OCT-2000	2000US-0241808.00
PR	20-OCT-2000	2000US-0241809.00
PR	20-OCT-2000	2000US-0241816.00
PR	01-NOV-2000	2000US-0244617.00
PR	08-NOV-2000	2000US-0246474.00
PR	08-NOV-2000	2000US-0246475.00
PR	08-NOV-2000	2000US-0246476.00
PR	08-NOV-2000	2000US-0246477.00
PR	08-NOV-2000	2000US-0246478.00
PR	08-NOV-2000	2000US-0246554.00
PR	08-NOV-2000	2000US-0246555.00
PR	08-NOV-2000	2000US-0246556.00
PR	08-NOV-2000	2000US-0246557.00
PR	08-NOV-2000	2000US-0246558.00
PR	08-NOV-2000	2000US-0246562.00
PR	08-NOV-2000	2000US-0246609.00
PR	08-NOV-2000	2000US-0246610.00
PR	08-NOV-2000	2000US-0246611.00
PR	11-NOV-2000	2000US-0246613.00
PR	11-NOV-2000	2000US-0249207.00
PR	17-NOV-2000	2000US-0249208.00
PR	17-NOV-2000	2000US-0249209.00
PR	17-NOV-2000	2000US-0249210.00
PR	17-NOV-2000	2000US-0249211.00
PR	17-NOV-2000	2000US-0249212.00

AC AAF21725;
 XX
 XX 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 112.
 XX
 XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO20005173-A1.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05881.
 XX
 XX 12-MAR-1999; 99US-0124270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM;
 XX WPI: 2000-611515/58.
 XX P-PSDB: AAB58822.
 XX
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 1: Page 566; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 S0 Sequence 743 BP; 150 A; 192 C; 238 G; 149 T; 14 other;

alignment_scores:
 Quality: 722.50 Length: 279
 Ratio: 3.992 Gaps: 3
 Percent Similarity: 64.875 Percent Identity: 63.441

alignment_block:
 US-09-866-034-2 x AAF21725 ..

Align seg 1/1 to: AAF21725 from: 1 to: 743

1 MetHisLysAlaGlyLeuGlyLeuGlyCysAlaArgAla.TTPAServ 17

||||| ||| ||||||| ||||| ... |||||
 16 ATGCACATGGCCAGGCTGTAGGCTTTGCT...GGCGNGAGATGG 62
 17 alArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 33
 63 TGGGATGGCCAGCTCCAGATGACCCCGGAGCCGCTACAAATAG 112
 34 ValAlaLeuValThrAlaSerThrAspGlyTleGlyPheAlaIleAla 50
 113 GTGGCCCTGTAAACGGCTCCACCGAGCGGATCGCTTCCG.ATCGCCG 161
 50 gArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgLysGln 67
 162 GCGTTGGCCGAGGACAG.GGCCAGTGTGTCTACGACGCGGAGACAGC 210
 67 InAsnValAspGlnAlaValAlaThrLeuGlnGlyGlnGlyLeuSerVal 83
 211 AGATGTGGACCGAGCGGTGGC.ACGCTGCARGGGAGGGCTGAGCGTG 259
 84 ThrGlyThrValCysHisValGlyLysAlaGlnAspArgGluArgLeu 100
 260 ACGGACACCTGTCANTGNTGGAGAGCGGAGACCGGAGCGCGTGT 309
 100 lAlaThrAlaValLysLeuHisGlyGlyTleAspIleLeuValSerSna 117
 310 GGCACGCGCTGTGAAGCTCATGGAGGTATCGATCTCTCCATG 359
 117 lAlaValAsnProPheGlySerIleMetAspValThrGlnGluVal 133
 360 CTGCTGTCAACCTTCTTGGAAAGCATATGATGATGATGAGAGAGTG 409
 134 TrpAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThr 150
 410 TGGGACAG... 418
 150 sAlaValValProLueMetGluLysArgGlyGlySerValIleVal 167
 418 418
 167 alSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsn 183
 418 418
 184 ValSerLysThrAlaLeuGlyLeuThrLysThrLeuAlaIleGlu 200
 418 418
 200 uAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLys 217
 418 418
 217 hrSerPheSerArgMetLeuThrMetAspLysGlnLysGlnIleSerMet 233
 419CTGTGATGGACAGAAAGGAAAGGAAAGCATG 451
 234 LysGluThrLeuArgIleArgArgLeuGlyGlnProLysPcysAlaG 250
 452 AAGAAACCTCGCGATTAAGAGATTAGCGAGACAGAGATTCTGCTG 501
 250 ylleValSerPheLeuGlySerGlnAspAlaSerTyrIleThrGlyGlu 267
 502 CATCGTGTCTTCTCTGTGTGATGAGATGCCACTACATCATCGGGAAA 551
 267 hrValValValGlyGlyGlyThrProSerArgLeu 278
 552 CAGT 586

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL16757

seq_documentation_block:

ID ABL16757 standard; DNA: 1022 BP.

AC ABL16757;

```

419 ACTGACCTTGAATGTCACCGCTTGAGAGTGGCATGTGACCGAACCCGAG 468
      :::::|||||::|  |||  |||||::|::|::|::|::|::|::|::|
419 ACTGACCTTGAATGTCACCGCTTGAGAGTGGCATGTGACCGAACCCGAG 468
      :::::|||||::|  |||  |||||::|::|::|::|::|::|::|::|
95 sparggluatygluvalalalathalavalylsleuhtsilylylleasp 111
      |||||::|::|::|  :::::||||  |||  :::::
469 ATCGCAACACGCTTTTGAGGAGGACCAATTACCAAGTTTGGCAAGCTCAAC 518
      |||||::|::|::|  :::::||||  |||  :::::
112 lleuvalserasnalalavalasiprothehegelyserlleuets 128
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
519 ATTCTGGTCCACATGACGACCCACCAATCTGGCTGGCGGTGTCCTGA 568
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
128 pvalthrglucluvaltyrpsaplysthrleuaspilleasnavillyslap 145
      :  |||||::|::|::|  :::::||||  :::::|||||::|::|::|
569 GTGCGATGAGAAGGTGTGGGACCAAGATCTTCGATGTGAACTGAAGACTT 618
      |||||::|::|::|  :::::||||  :::::|||||::|::|::|
145 roalaleumecthrlysalavalalproglumetglulsarysilylyl 161
      |||||::|::|::|  :::::||||  :::::|||||::|::|::|
619 CCTATCTGCGGCCAAGAGGACCTGCTCTCCGCGCCACCAACAAAGAAC 668
      |||||::|::|::|  :::::||||  :::::|||||::|::|::|
162 glyservalvallevalserseillaalapheserproserprogl 178
      :::::|||||::|::|::|  :::::||||  :::::|||||::|::|::|
669 TCCAGCATGCTTTTCGCTCTCTCCATTCGTGGCTATGATGCTTGAGCT 718
      :::::|||||::|::|::|  :::::||||  :::::|||||::|::|::|
178 ypheserprotyrasnavalserlysthralaleuleuglydeuthrlyst 195
      :::::||||  |||||::|::|::|  |||||::|::|::|
719 ACTGGAGGCTATTTCGTCGACCAAGACCGGCGTATTGGCTTGACCAAG 768
      :::::||||  |||||::|::|::|  |||||::|::|::|
195 hrleumalalleuileualalproargasulleargvalasncysleuala 211
      ::  ::  |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
769 CACCCGCGCAAGATCTGGCGCCGACGAGGCAATCCGCTCACTGCTGCTGCT 818
      ::  ::  |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
212 proglileuilelysthrserpheserfargmetleuutrrhmetasplysli 228
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
819 CCAGGAGTATCATGACCAAGTCTCTCAAAAGCACTGTACAGCAAGATGAGTC 868
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
228 ulysgluclusermetclyscluthrleuarqileargargleuglyslup 245
      :  :::::||||  :::::||||  :::::|||||::|::|::|
869 GGGGAATGAAAGCGCTTWTGAGCAAAATATCCCAATGCGTCTGGCAACA 918
      :  :::::||||  :::::||||  :::::|||||::|::|::|
245 rogluaspysalaglyllevalserpheleucyssergluaspalasr 261
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
919 GTGAGGAGATGCTGGCGCTGCTCCCTTTTGGCTCCGAGGATGCGCGC 968
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
262 tyfletthrlygluuthrvalvalalgllygllythrproserargyle 278
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
969 TACATTAACCGAGAGATTCATTGTGGCTGGCGGTGGATGACAGCTCGTTT 1018
      |||||::|::|::|  |||||::|::|::|  |||||::|::|::|
278 u 278
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1019 G 1019

```

seq_documentation_block:	
ID	ABL16756 standard; DNA; 3147 BP.
XX	
AC	ABL16756;
XX	
DT	26-MAR-2002 (first entry).
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 1741.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS
 Claim 1: SEQ ID NO 1741; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB161175), expressed DNA
 CC sequences (AB101840-AB161175) and the encoded proteins
 CC (AB157737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 3147 BP; 884 A; 785 C; 726 G; 752 T; 0 other;
 alignment_scores:
 Quality: 619.00 Length: 309
 Ratio: 2.879 Gaps: 3
 Percent Similarity: 69.579 Percent Identity: 44.984
 alignment_block:
 US-09-866-034-2 x AB16756/rev ..
 Align seg 1/1 to reverse of: AB16756 from: 1 to: 3147
 12 ArgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrArgArgAs 28
 1923 CGCTTGTCCAGCAGTAGTCAAGTTCACAGCCGCCGACATGAAGCT.. 1876
 28 pProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGly... 43
 1875 ...TTGGCAGGAAAGTAGCCGTGGTCACCGCTCCACGATGGGTAG 1830
 43 43
 1829 CAGCTCATAGAAACTGGAGAGATCGCACCCCTAATTCATTAATC 1780
 44 116glyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAl 57
 1779 CCTCCAGCATGTGTTCCGATTGCCAAAGCGTCGCCCAAGATGGCG 1730
 57 aHisValValSerSerArgLysGlnGlnAsnValAspGlnAlaValA 74
 1729 CCGAGGTGTCACGATCGCAAGCAAAAGATGTGACTGTGCTGTG 1680
 74 laThrLeuGlnGlyGlyLeuSerValThrGlyThrValLysHisVal 90
 1679 CGAGCTGTGCAAACTGAATTAATGTCCACGCTCTGAAGTCCCATGTC 1630
 91 GlyLysAlaGlnAspArgLysArgLeuValAlaThrAlaValLysLeu 107
 1629 AGCGAAGCCGAGATCGCAAAAGCTCTTTGAGAGACACATTAGCAAGTT 1580
 107 sGlyGlyLysPheLeuValSerAsnAlaValAsnProPhePheG 124
 1579 TGGCAAGCTCAATCTGTGTGTCAGCAATGACCCCAATCTCGCGTGC 1530
 124 LysIleMetAspValThrGluGluValThrAspLysThrLeuAspIle 140

1529 GCGGTGCTCTCGAGTCGATGAGAGAGCTGTGGGACAAAGATCTTCGATG 1480
 141 AsnValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetG 157
 1479 AACGTGAAGAGTCTCTATCTGTGCGCAAGAGGAGCTGCTCTGCG 1430
 157 uLysArgGlyGlySerValValIleValSerSerIleAlaAlaPhe 174
 1429 CCAGCAAAAGAACTCCACATCGTTTCTGCTCCCTCCATCTGCGTATG 1380
 174 erProSerProGlyPheSerPro..... 181
 1379 ATGCCCTTGA..GGTTAGTCGCTTAATCTCGGATTCAGTTTCATGCTT 1331
 182TyrAsnValSerLys 186
 1330 ATTCGCCGTACTTAATTCGATAGTACTGAGAGCTTAATTCGACGAA 1281
 186 sThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaPro 203
 1280 GACCGCGCTGANTGCTTGACCAAGGACGCCGCAAGAGATGCGCGCG 1231
 203 rGAsnIleArgValAsnLysLeuAlaProGlyLeuIleLysThrSerPhe 219
 1230 AGGCAATTCGCGTCAACCTGCTGCTCCAGAGATCATCAGCAAAAGTTC 1181
 220 SerArgMetLeuTrpMetAspLysGluLysGluGluSerMetLysGluTh 236
 1180 TCGAAAGCAGCTGTACGAAATGAGTCGGCAATGAGCGCTTGACCAA 1131
 236 rLeuArgIleArgArgLeuGlyLysGluProGluAspCysAlaGlyLys 253
 1130 AATACCATGCGCTGCTGCGGACACAGAGAGATGCTGCGTGTCT 1081
 253 erPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrValVal 269
 1080 CTTTGTGCTCGAGATGCGGCTACATTACGAGAGATGATCATGTTG 1031
 270 ValGlyGlyGlyThrProSerArgLeu 278
 1030 GCTGCGGTGGAATGACAGCCTTTG 1004
 seq_name: /STDS1/gcgdata/geneseq/geneseg-emb1/NN2001B.DAT.AB107994
 seq_documentation_block:
 ID AB107994 standard; cDNA; 2699 BP.
 XX
 AC AB107994;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18464.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PE 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.

2254 TCCTATCTGCTGGCCAAGAGGACACTGCCTCTCCTGGCCAGCAAAAGAA 2205

21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510

```

seq_name: /SIDs1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.AAC46747
seq_documentation_block:
ID AAC46747 standard; DNA; 1123 BP.
XX
XX AAC46747;
XX
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51254.
XX
XX Hybridisation assay; genetic mapping; gene expression control
XX protein-identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
XX
XX 06-SEP-2000.
XX
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.

```

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

Align seg 1/1 to: AAC46747 from: 1 to: 1123

1. The first part of the document is a list of references. The references are listed in a standard format, with the author's name, the title of the work, and the publisher. The references are as follows:

1. J. H. Van Vleck, *Phys. Rev.* **32**, 47 (1928).
2. J. H. Van Vleck, *Phys. Rev.* **33**, 47 (1929).
3. J. H. Van Vleck, *Phys. Rev.* **34**, 47 (1930).
4. J. H. Van Vleck, *Phys. Rev.* **35**, 47 (1931).
5. J. H. Van Vleck, *Phys. Rev.* **36**, 47 (1932).
6. J. H. Van Vleck, *Phys. Rev.* **37**, 47 (1933).
7. J. H. Van Vleck, *Phys. Rev.* **38**, 47 (1934).
8. J. H. Van Vleck, *Phys. Rev.* **39**, 47 (1935).
9. J. H. Van Vleck, *Phys. Rev.* **40**, 47 (1936).
10. J. H. Van Vleck, *Phys. Rev.* **41**, 47 (1937).
11. J. H. Van Vleck, *Phys. Rev.* **42**, 47 (1938).
12. J. H. Van Vleck, *Phys. Rev.* **43**, 47 (1939).
13. J. H. Van Vleck, *Phys. Rev.* **44**, 47 (1940).
14. J. H. Van Vleck, *Phys. Rev.* **45**, 47 (1941).
15. J. H. Van Vleck, *Phys. Rev.* **46**, 47 (1942).
16. J. H. Van Vleck, *Phys. Rev.* **47**, 47 (1943).
17. J. H. Van Vleck, *Phys. Rev.* **48**, 47 (1944).
18. J. H. Van Vleck, *Phys. Rev.* **49**, 47 (1945).
19. J. H. Van Vleck, *Phys. Rev.* **50**, 47 (1946).
20. J. H. Van Vleck, *Phys. Rev.* **51**, 47 (1947).
21. J. H. Van Vleck, *Phys. Rev.* **52**, 47 (1948).
22. J. H. Van Vleck, *Phys. Rev.* **53**, 47 (1949).
23. J. H. Van Vleck, *Phys. Rev.* **54**, 47 (1950).
24. J. H. Van Vleck, *Phys. Rev.* **55**, 47 (1951).
25. J. H. Van Vleck, *Phys. Rev.* **56**, 47 (1952).
26. J. H. Van Vleck, *Phys. Rev.* **57**, 47 (1953).
27. J. H. Van Vleck, *Phys. Rev.* **58**, 47 (1954).
28. J. H. Van Vleck, *Phys. Rev.* **59**, 47 (1955).
29. J. H. Van Vleck, *Phys. Rev.* **60**, 47 (1956).
30. J. H. Van Vleck, *Phys. Rev.* **61**, 47 (1957).
31. J. H. Van Vleck, *Phys. Rev.* **62**, 47 (1958).
32. J. H. Van Vleck, *Phys. Rev.* **63**, 47 (1959).
33. J. H. Van Vleck, *Phys. Rev.* **64**, 47 (1960).
34. J. H. Van Vleck, *Phys. Rev.* **65**, 47 (1961).
35. J. H. Van Vleck, *Phys. Rev.* **66**, 47 (1962).
36. J. H. Van Vleck, *Phys. Rev.* **67**, 47 (1963).
37. J. H. Van Vleck, *Phys. Rev.* **68**, 47 (1964).
38. J. H. Van Vleck, *Phys. Rev.* **69**, 47 (1965).
39. J. H. Van Vleck, *Phys. Rev.* **70**, 47 (1966).
40. J. H. Van Vleck, *Phys. Rev.* **71**, 47 (1967).
41. J. H. Van Vleck, *Phys. Rev.* **72**, 47 (1968).
42. J. H. Van Vleck, *Phys. Rev.* **73**, 47 (1969).
43. J. H. Van Vleck, *Phys. Rev.* **74**, 47 (1970).
44. J. H. Van Vleck, *Phys. Rev.* **75**, 47 (1971).
45. J. H. Van Vleck, *Phys. Rev.* **76**, 47 (1972).
46. J. H. Van Vleck, *Phys. Rev.* **77**, 47 (1973).
47. J. H. Van Vleck, *Phys. Rev.* **78**, 47 (1974).
48. J. H. Van Vleck, *Phys. Rev.* **79**, 47 (1975).
49. J. H. Van Vleck, *Phys. Rev.* **80**, 47 (1976).
50. J. H. Van Vleck, *Phys. Rev.* **81**, 47 (1977).
51. J. H. Van Vleck, *Phys. Rev.* **82**, 47 (1978).
52. J. H. Van Vleck, *Phys. Rev.* **83**, 47 (1979).
53. J. H. Van Vleck, *Phys. Rev.* **84**, 47 (1980).
54. J. H. Van Vleck, *Phys. Rev.* **85**, 47 (1981).
55. J. H. Van Vleck, *Phys. Rev.* **86**, 47 (1982).
56. J. H. Van Vleck, *Phys. Rev.* **87**, 47 (1983).
57. J. H. Van Vleck, *Phys. Rev.* **88**, 47 (1984).
58. J. H. Van Vleck, *Phys. Rev.* **89**, 47 (1985).
59. J. H. Van Vleck, *Phys. Rev.* **90**, 47 (1986).
60. J. H. Van Vleck, *Phys. Rev.* **91**, 47 (1987).
61. J. H. Van Vleck, *Phys. Rev.* **92**, 47 (1988).
62. J. H. Van Vleck, *Phys. Rev.* **93**, 47 (1989).
63. J. H. Van Vleck, *Phys. Rev.* **94**, 47 (1990).
64. J. H. Van Vleck, *Phys. Rev.* **95**, 47 (1991).
65. J. H. Van Vleck, *Phys. Rev.* **96**, 47 (1992).
66. J. H. Van Vleck, *Phys. Rev.* **97**, 47 (1993).
67. J. H. Van Vleck, *Phys. Rev.* **98**, 47 (1994).
68. J. H. Van Vleck, *Phys. Rev.* **99**, 47 (1995).
69. J. H. Van Vleck, *Phys. Rev.* **100**, 47 (1996).
70. J. H. Van Vleck, *Phys. Rev.* **101**, 47 (1997).
71. J. H. Van Vleck, *Phys. Rev.* **102**, 47 (1998).
72. J. H. Van Vleck, *Phys. Rev.* **103**, 47 (1999).
73. J. H. Van Vleck, *Phys. Rev.* **104**, 47 (2000).
74. J. H. Van Vleck, *Phys. Rev.* **105**, 47 (2001).
75. J. H. Van Vleck, *Phys. Rev.* **106**, 47 (2002).
76. J. H. Van Vleck, *Phys. Rev.* **107**, 47 (2003).
77. J. H. Van Vleck, *Phys. Rev.* **108**, 47 (2004).
78. J. H. Van Vleck, *Phys. Rev.* **109**, 47 (2005).
79. J. H. Van Vleck, *Phys. Rev.* **110**, 47 (2006).
80. J. H. Van Vleck, *Phys. Rev.* **111**, 47 (2007).
81. J. H. Van Vleck, *Phys. Rev.* **112**, 47 (2008).
82. J. H. Van Vleck, *Phys. Rev.* **113**, 47 (2009).
83. J. H. Van Vleck, *Phys. Rev.* **114**, 47 (2010).
84. J. H. Van Vleck, *Phys. Rev.* **115**, 47 (2011).
85. J. H. Van Vleck, *Phys. Rev.* **116**, 47 (2012).
86. J. H. Van Vleck, *Phys. Rev.* **117**, 47 (2013).
87. J. H. Van Vleck, *Phys. Rev.* **118**, 47 (2014).
88. J. H. Van Vleck, *Phys. Rev.* **119**, 47 (2015).
89. J. H. Van Vleck, *Phys. Rev.* **120**, 47 (2016).
90. J. H. Van Vleck, *Phys. Rev.* **121**, 47 (2017).
91. J. H. Van Vleck, *Phys. Rev.* **122**, 47 (2018).
92. J. H. Van Vleck, *Phys. Rev.* **123**, 47 (2019).
93. J. H. Van Vleck, *Phys. Rev.* **124**, 47 (2020).
94. J. H. Van Vleck, *Phys. Rev.* **125**, 47 (2021).
95. J. H. Van Vleck, *Phys. Rev.* **126**, 47 (2022).
96. J. H. Van Vleck, *Phys. Rev.* **127**

missing at time of publication, meaning no sequences are present for

[illegible]